

AD \_\_\_\_\_

Award Number: DAMD17-98-1-8042

TITLE: New Classes of Conditional Toxins as Therapeutic Agents  
Against Breast Cancer

PRINCIPAL INVESTIGATOR: Alexander Varshavsky, Ph.D.

CONTRACTING ORGANIZATION: California Institute of Technology  
Pasadena, California 91125

REPORT DATE: April 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are  
those of the author(s) and should not be construed as an official  
Department of the Army position, policy or decision unless so  
designated by other documentation.

DATA CHECKED AND APPROVED 4

20010124 092

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

<b>1. AGENCY USE ONLY (Leave blank)</b>	<b>2. REPORT DATE</b> April 2000	<b>3. REPORT TYPE AND DATES COVERED</b> Annual (1 Apr 99 - 31 Mar 00)	
<b>4. TITLE AND SUBTITLE</b> New Classes of Conditional Toxins as Therapeutic Agents Against Breast Cancer		<b>5. FUNDING NUMBERS</b> DAMD17-98-1-8042	
<b>6. AUTHOR(S)</b> Alexander Varshavsky, Ph.D.			
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> California Institute of Technology Pasadena, California 91125		<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>E-MAIL:</b> avarsh@cco.caltech.edu			
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		<b>10. SPONSORING / MONITORING AGENCY REPORT NUMBER</b>	
<b>11. SUPPLEMENTARY NOTES</b> This report contains colored photographs			
<b>12a. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for public release; distribution unlimited		<b>12b. DISTRIBUTION CODE</b>	
<b>13. ABSTRACT (Maximum 200 Words)</b>			
<p>During the second year of support by the grant DAMD17-98-1-8042 (The Idea Grant) we focused on the following projects:</p> <ol style="list-style-type: none"> <li>1) Further development of signal-regulated, cleavage-mediated toxins. During this year, we completed the work on the CUP9-mediated, HIV protease-dependent conditional toxins. A paper describing these results is nearly finished, and will be submitted for publication this summer.</li> <li>2) The effort of the previous two years to construct better portable degrons (degradation signals) that can be used, in particular, for designing of conditional toxins was successful. A paper describing these results was published at the end of 1999 (Suzuki &amp; Varshavsky. <i>Degradation signals in the lysine-asparagine sequence space</i>. <b>EMBO J.</b> 18:6017-6026, 1999).</li> <li>3) In 1999, we completed the construction of a bivalent inhibitor of the N-end rule pathway, a useful reagent for conditional-toxin projects, and published this new approach (Kwon &amp; Varshavsky. <i>Bivalent inhibitor of the N-end rule pathway</i>. <b>J. Biol. Chem.</b> 274:18135-18139, 1999).</li> <li>4) Other projects that encompass the subject of this grant are under way. They are briefly described below.</li> </ol>			
<b>14. SUBJECT TERMS</b> Breast Cancer proteolysis / degradation signals / codominance / conditional toxins / cancer		<b>15. NUMBER OF PAGES</b> 241	
		<b>16. PRICE CODE</b>	
<b>17. SECURITY CLASSIFICATION OF REPORT</b> Unclassified	<b>18. SECURITY CLASSIFICATION OF THIS PAGE</b> Unclassified	<b>19. SECURITY CLASSIFICATION OF ABSTRACT</b> Unclassified	<b>20. LIMITATION OF ABSTRACT</b> Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

Where copyrighted material is quoted, permission has been obtained to use such material.

Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

  
4/30/00  
PI - Signature Date

## **Table of Contents**

Report Documentation Page & Abstract	page 2
Foreword	Page 3
<b>Progress Report for the DOD Grant DAMD17-98-1-8042</b>	<b>page 1-4</b>
Curriculum Vitae of PI (Alexander Varshavsky)	page 1-10
Reprints and preprints of the recent paper by the laboratory (three sets)	

## **Progress Report for the DOD Grant DAMD 17-98-1-8042**

April 1999-April 2000 (second year of support)

During the second year of support by the grant DAMD17-98-1-8042 (the Idea Grant) we completed several projects described last year, and continued with several other projects.

### **Signal-regulated, cleavage-mediated toxins**

During 1999 and early 2000, we completed the work on the first example of a cleavage-mediated conditional toxin. A draft of the paper about this class of conditional toxins, to be submitted for publication soon (it is being revised as I write) is enclosed. (Davydov & Varshavsky. *Conditional toxin against yeast cells expressing HIV protease*. **J. Biol. Chem.** (to be submitted in summer; early preprint enclosed).

### **Construction of degradation signals in the lysine-asparagine sequence space**

Among the problems to be solved on the way to multitarget conditional toxins is the paucity of strong, portable N-degrons that are free of drawbacks of the existing degrons of this class. These drawbacks include the tendency of proteins bearing such degrons to undergo endoproteolytic cleavages that sever the N-degron from the rest of the protein, and thereby contribute to the high background of assays with N-degron-bearing proteins. As a part of our ongoing effort to construct better degradation signals, we developed a new approach, which makes it possible to identify and isolate specific degradation signals in the sequence space of just two amino acids, lysine and asparagine A paper describing these results was published in late 1999. (Suzuki & Varshavsky. *Degradation signals in the lysine-asparagine sequence space*. **EMBO J.** 18:6017-6026, 1999.) (reprint enclosed.)

### **Bivalent Inhibitor of Protein Degradation by the N-End Rule Pathway**

In a project relevant to the theme of this grant, we have developed a specific bivalent inhibitor of the N-end rule pathway. The relevance of this advance stems from the fact the bivalent inhibitor (described below) improved the understanding of the targeting (substrate-binding) sites of N-recognin, the E3 component of the ubiquitin-dependent N-end rule pathway,

which is at the center of our sitoxins and comtoxin projects. Eventually, the designs of conditional toxins should be able to involve any of the several ubiquitin-dependent proteolytic pathways; the current focus on the N-end rule pathway reflects not only the interest of this laboratory but also the fact that this pathway is particularly well understood mechanistically. Moreover, in staying within the N-end rule pathway, it is possible to use the portable, highly active N-degrons, an advantage that is not available with most of the other ubiquitin-dependent pathways. A paper describing these results was published in 1999. (Kwon & Varshavsky. *Bivalent inhibitor of the N-end rule pathway*. **J. Biol. Chem.** 274:18135-18139, 1999.) (reprint enclosed.)

### **Some of the Projects in Progress**

#### **Construction and Use of the Library Expressing Random Fragments of the Green Fluorescent Protein (GFP) and Src Protein Kinase**

The aim of this ongoing project is to develop a generally applicable method for interfering with the *in vivo* function of a target protein of interest through the expression of a specific fragment of the target that would bind to a conformationally immature (nascent) target protein and thereby either arrest or strongly retard its conformational maturation. As a result, the complex of the target and its self-peptide would be nonfunctional insofar as the target's function is concerned. In addition, the interfering peptide would be equipped with a strong N-terminal degradation signal (the N-degron, previously studied by this laboratory). A permanently or transiently stable complex between a target protein and its self-peptide bearing an N-degron would be targeted for processive degradation by the N-end rule pathway, one of the pathways of the ubiquitin system. Once developed, this method could be applied to effect selective functional inhibition of undesirable proteins such as, for example, overexpressed oncoproteins. This trans-degradation method would also become a part of the comtoxin approach described in the DOD Idea Grant that supports these studies. To develop the trans-degradation technique, we chose Green Fluorescent Protein (GFP) as a model target protein. (Our next, clinically relevant, target is Src, a protein tyrosine kinase and an proto-oncogene activated in a number of human cancers, including breast cancer.) We carried out the construction of a library of clones which expressed polypeptide fragments derived from random segments of GFP. These fragments were

expressed in the yeast *S. cerevisiae* from the P<sub>GAL</sub> promoter as linear fusions to ubiquitin. As shown previously by this laboratory, the N-terminal ubiquitin moiety of the fusion is cotranslationally cleaved off by ubiquitin-specific proteases, making it possible to expose any desired amino acid residue at the N-terminus of the resulting polypeptide fragment. In developing this approach, we are proceeding in steps, at first constructing a ubiquitin fusion-based library of GFP fragments as such. The second-generation library will have the same fragments bearing the 40-residue N-terminal extension called Arg-e<sup>ΔK</sup>, which contains the above-described N-degron. The library expressing random GFP fragments was constructed using DNase I in the presence of Mn<sup>2+</sup> ions to fragment the GFP cDNA, and specific oligonucleotide adapters to insert the resulting fragments into the expression vector. More recently, we began the construction of analogous libraries with v-Src and c-Src trans-targeting, degron-containing reporter. The v-Src and c-Src cDNAs were obtained from Dr. D. Morgan (Univ. of California, Berkeley). They were tested by sequencing the 5'- and 3'-ends (~200 bp each) of the coding regions, and confirming that the resulting sequences matched those for v-Src and c-Src in the GenBank database. Restriction fragments containing the Src genes were isolated by agarose gel electrophoresis, and are being used to construct random-fragment libraries as described above. Taking advantage of the unique Nco I site at the 5'-end the v-Src cDNA, the v-Src coding region (as a filled-in Nco I-Xho I fragment) was inserted into the integration vector pIP123-CUPR. This placed the Src ORF under the control of the copper-inducible P<sub>CUP1</sub> promoter. It remains to be determined, in this ongoing project, whether this promoter is sufficiently tight (low enough level of Src in the uninduced state) and sufficiently strong (high enough level of Src in the induced state to cause growth arrest). (In contrast to the GFP-based library, where the screen is being carried out by fluorescence-activated cell sorting (FACS), our readout and the screen with Src are based on the known toxicity of mammalian Src to the yeast *S. cerevisiae*.) Examination of the three-dimensional structures of Src family tyrosine kinases shows that a loop connecting the SH2 domain to the kinase domain is bound to the SH3 domain of Src and sandwiched between the SH3 domain and the N-terminal half of the kinase when Src is in the inactive state. The kinase-inhibiting interaction requires phosphorylation of Src at its C-terminal region, itself bound to the SH2 domain. This area is a promising candidate for a trans-targeting N-degron. Our aims in this set of projects are, first, to optimize the library-based

interference approach described above, using GFP as both the target and reporter, then to expand this approach to the Src kinase, and to link an N-degron to the interfering fragments identified in both the GFP and Src library screen. This multistep strategy should result in a generally applicable method for metabolically destabilizing (and thereby functionally inhibiting) any protein of interest, including the targets that must be down-regulated for either arresting or selectively killing the breast adenocarcinoma cells.

Over the last year, this project ran into technical difficulties, but we are committed to it, and are continuing this work.

### **Designing Asparagine-Lysine Degrons That Work in Mammalian Cells**

We are beginning to test the lysine-asparagine N-degrons of *S. cerevisiae* identified in the preceding work supported by this grant (Suzuki & Varshavsky. *Degradation signals in the lysine-asparagine sequence space*. **EMBO J.** 18:6017-6026, 1999) (reprint enclosed) in mammalian cells. Our aim is to determine whether any of the highly active yeast N-degrons of this class that have been identified in yeast are comparably powerful in mammals. If they are, several applications, including those that encompass conditional toxins will become possible right away. If these degrons are not as strong in the mammalian-cell setting, we will modify them, using techniques that have already proved successful in yeast, to produce a set of degrons that would confer on a mammalian protein reporter an *in vivo* half-life of 3 minutes or less. With degrons of this quality, one could design proteolysis-based conditional mutants in mice, and conditional toxins as well. This project is under way.

Enclosed with this Progress Report are the recently published papers (1999-2000; three sets), and three copies of the PI's CV. The Progress Report describes projects that were supported, either partially or entirely, by the DOD Idea Grant during its second year.

April 2000

## CURRICULUM VITAE

<b>Name</b>	Alexander J. Varshavsky
<b>Date and Place of Birth</b>	November 8, 1946, Moscow, Russia
<b>Citizenship</b>	U.S. citizen.
<b>Address</b>	Division of Biology, 147-75 California Institute of Technology 1200 East California Boulevard Pasadena, CA 91125
<b>Telephone</b>	626-395-3785 (office); 818-541-9791 (home)
<b>Fax</b>	626-440-9821 (office fax); 818-248-5245 (home fax)
<b>Email</b>	avarsh@caltech.edu

### Academic Appointments and Education

1970	B. S. in Chemistry, Department of Chemistry, Moscow University, Moscow, Russia
1973	Ph. D. in Biochemistry, Institute of Molecular Biology, Moscow, Russia
1973-1976	Research Fellow, Institute of Molecular Biology, Moscow, Russia
1977-1980	Assistant Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts
1980-1986	Associate Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts
1986-1992	Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts
1992-present	Howard and Gwen Laurie Smits Professor of Cell Biology, Division of Biology, California Institute of Technology, Pasadena, California

### Professional Service

1983-1987	Member, Molecular Cytology Study Section, NIH
1993	Co-organizer, the 1993 Banbury Conference on the Ubiquitin System, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

## Honors

1987	Elected to the American Academy of Arts and Sciences
1995	Elected to the National Academy of Sciences
1998	Merit Award, by the National Institutes of Health
1998	Novartis-Drew Award, by Novartis, Inc. and the Drew University
1999	Gairdner International Award, by the Gairdner Foundation, Canada (with A. Hershko)
2000	Shubitz Prize in Cancer Research, by the University of Chicago
2000	Hoppe-Seyler Award, by the German Biochemical Society
2000	Sloan Prize, by the General Motors Cancer Research Foundation (with A. Hershko)

## Selected Publications (1968-present)

(grouped by the fields; numbered chronologically)

### Chromosome Structure and Gene Expression

1. Varshavsky, A. Regulation of synthesis of genetic repressors in bacteria. **Mol. Biol.** (Russia) 2:13-20 (1968).
4. Ilyin, Y. V., Varshavsky, A., Mickelsaar, U. N. and Georgiev, G. P. Redistribution of proteins in mixtures of nucleoproteins, DNA and RNA. **Eur. J. Biochem.** 22:235-245 (1971).
5. Varshavsky, A. and Georgiev, G. P. Clustered arrangement of histones F2a1 and F3 in chromosomal deoxyribonucleoproteins. **Biochim. Biophys. Acta** 281:449-674 (1972).
9. Varshavsky, A. and Georgiev, G. P. Redistribution of histones during unfolding of chromosomal DNA. **Mol. Biol. Reports** 1:143-148 (1973).
12. Varshavsky, A. and Ilyin, Y. V. Salt treatment of chromatin induces redistribution of histones. **Biochim. Biophys. Acta** 340:207-217 (1974).
14. Ilyin, Y. V., Bayev, A. A. Jr., Zhuse, A. L. and Varshavsky, A. Histone-histone proximity in chromatin as revealed by imidoester crosslinking. **Mol. Biol. Reports** 1:343-348 (1974).
17. Varshavsky, A. and Bakayev, V. V. Nu-bodies and free DNA in chromatin lacking histone H1. **Mol. Biol. Reports** 2:209-217 (1975).
23. Varshavsky, A., Bakayev, V. V., Chumakov, P. M. and Georgiev, G. P. Minichromosome of simian virus 40: presence of histone H1. **Nucl. Acids Res.** 3:2101-2114 (1976).
27. Varshavsky, A. Structural and functional organization of eukaryotic chromosomes. **Biol. Zentralblatt** 95:301-316 (1976).

30. Bakayev, V. V., Bakayeva, T. G. and Varshavsky, A. Nucleosomes and subnucleosomes: heterogeneity and composition. **Cell** 11:619-630 (1977).
31. Varshavsky, A., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. and Georgiev, G. P. Histone-like proteins in the *E. coli* chromosome. **Nucl. Acids Res.** 4:2725-2745 (1977).
33. Varshavsky, A., Bakayev, V. V., Nedospasov, S. A. and Georgiev, G. P. On the structure of eukaryotic, prokaryotic and viral chromatin. **Cold Spring Harbor Symp. Quant. Biol.** 42:457-472 (1977).
37. Varshavsky, A., Sundin, O. and Bohn, M. SV40 viral minichromosome: preferential exposure of the origin of replication. **Nucl. Acids Res.** 5:3469-3478 (1978).
38. Varshavsky, A., Sundin, O. and Bohn, M. A 400 bp stretch of SV40 viral DNA that includes the origin of replication is exposed in SV40 minichromosomes. **Cell** 16:453-466(1979).
39. Sundin, O. and Varshavsky, A. Staphylococcal nuclease makes a single nonrandom cut in the SV40 viral minichromosome. **J. Mol. Biol.** 132:535-546 (1979).
45. Levinger, L. and Varshavsky, A. Drosophila heat shock proteins are associated with nuclease-resistant, high salt-resistant nuclear structures. **J. Cell Biol.** 90:793-796 (1981).
50. Barsoum, J., Levinger, L. and Varshavsky, A. On the chromatin structure of the amplified, transcriptionally active gene for dihydrofolate reductase in mouse cells. **J. Biol. Chem.** 257:5274-5282 (1982).
51. Levinger, L. and Varshavsky, A. Protein D1 preferentially binds AT-DNA and is a component of *Drosophila melanogaster* nucleosomes containing AT-rich satellite DNA. **Proc. Natl. Acad. Sci. USA** 79:7152-7156 (1982).
54. Wu, K., Strauss, F. and Varshavsky, A. Nucleosome arrangement in green monkey  $\alpha$ -satellite chromatin. **J. Mol. Biol.** 170:93-117 (1983).
67. Barsoum, J. and Varshavsky, A. Preferential localization of variant nucleosomes near the 5'-end of the mouse dihydrofolate reductase gene. **J. Biol. Chem.** 260:7688-7697 (1985).
73. Solomon, M. J., Strauss, F. and Varshavsky, A. A mammalian HMG protein recognizes a stretch of six AT base pairs in duplex DNA. **Proc. Natl. Acad. Sci. USA** 83:1276-1289 (1986).
80. Peck, L. J., Millstein, L., Eversole-Cire, P., Gottesfeld, J. M. and Varshavsky, A. Transcriptionally inactive oocyte-type 5S RNA genes of *Xenopus laevis* are complexed with TFIIIA *in vitro*. **Mol. Cell. Biol.** 7:3503-3510 (1987).
91. Winter, E. and Varshavsky, A. A DNA-binding protein that recognizes oligo dA-oligo dT tracts. **EMBO J.** 8:1867-1877 (1989).

### **DNA Replication, Gene Amplification, Transporters & Drug Resistance**

41. Sundin, O. and Varshavsky, A. Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. **Cell** 21:103-114 (1980).
43. Varshavsky, A. On the possibility of metabolic control of replicon "misfiring": relationship to emergence of malignant phenotypes in mammalian cell lineages. **Proc. Natl. Acad. Sci. USA** 78:3673-3677 (1981).
44. Varshavsky, A. Phorbol ester dramatically increases incidence of methotrexate-resistant cells: possible mechanisms and relevance to tumor promotion. **Cell** 25:561-572 (1981).

46. Sundin, O. and Varshavsky, A. Arrest of segregation leads to accumulation of highly intertwined catenated dimers: dissection of the final stages of SV40 DNA replication. **Cell** 25:659-669 (1981).
58. Varshavsky, A. Do stalled replication forks synthesize a specific alarmone? **J. Theoret. Biol.** 105:707-714, 1983.
55. Barsoum, J. and Varshavsky, A. Mitogenic hormones and tumor promoters greatly increase the incidence of cells bearing amplified dihydrofolate reductase genes. **Proc. Natl. Acad. Sci. USA** 80:5330-5334 (1983).
56. Varshavsky, A. Diadenosine 5', 5'''-P<sup>1</sup>,P<sup>4</sup>-tetraphosphate: a pleiotropically acting alarmone? **Cell** 34:711-712 (1983).
59. Snapka, R. and Varshavsky, A. Loss of unstably amplified dihydrofolate reductase genes from mouse cells is accelerated by hydroxyurea. **Proc. Natl. Acad. Sci. USA** 80:7533-7537 (1983).
63. Roninson, I., Abelson, H. T., Housman, D. E., Howell, N. and Varshavsky, A. Amplification of specific DNA sequences correlates with multidrug resistance in Chinese hamster cells. **Nature** 309:626-628 (1984).
71. Ciccarelli, R. B., Solomon, J. J., Varshavsky, A. and Lippard, S. J. *In vivo* effects of cis- and trans-diaminedichloroplatinum (II) on SV40 chromosomes: differential repair, DNA-protein crosslinking, and inhibition of replication. **Biochemistry** 24:7533-7540 (1985).
72. Gros, P., Croop, J., Roninson, I., Varshavsky, A. and Housman, D. E. Isolation and characterization of DNA sequences amplified in multidrug-resistant hamster cells. **Proc. Natl. Acad. Sci. USA** 83:337-341 (1986).
81. Solomon, M. J. and Varshavsky, A. A nuclease-hypersensitive region forms *de novo* after chromosome replication. **Mol. Cell. Biol.** 7:3822-3825 (1987).
93. McGrath, J. P. and Varshavsky, A. The yeast *STE6* gene encodes a homolog of the mammalian multidrug resistance P-glycoprotein. **Nature** 340:400-404 (1989).

## New Methods

16. Bakayev, V. V., Melnickov, A. A., Osicka, V. A. and Varshavsky, A. Isolation and characterization of chromatin subunits. **Nucl. Acids Res.** 2:1401-1419 (1975). (*Low ionic strength electrophoretic technique for separation of DNA-protein complexes.*)
22. Varshavsky, A., Bakayev, V. V. and Georgiev, G. P. Heterogeneity of chromatin subunits and location of histone H1. **Nucl. Acids Res.** 3:477-492 (1976). (*Fractionation of nucleosomes by the low ionic strength electrophoresis, the forerunner of gel shift assay.*)
42. Levinger, L., Barsoum, J. and Varshavsky, A. Two-dimensional hybridization mapping of nucleosomes. **J. Mol. Biol.** 146:287-304 (1981).
49. Boyce, F., Sundin, O., Barsoum, J. and Varshavsky, A. New way to isolate SV40 viral minichromosomes: use of a thiol-specific reagent. **J. Virol.** 42:292-296 (1982).
64. Strauss, F. and Varshavsky, A. A protein binds to a satellite DNA repeat at three sites which would be brought into proximity by DNA folding in the nucleosome. **Cell** 37:889-901 (1984). (*First use of the gel shift assay for detection and purification of specific DNA-binding proteins from cell extracts.*)

69. Solomon, M. J. and Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. **Proc. Natl. Acad. Sci. USA** 82:6470-6474 (1985).
74. Swerdlow, P. S., Finley, D. and Varshavsky, A. Enhancement of immunoblot sensitivity by heating of hydrated filters. **Analyt. Biochem.** 156:147-153 (1986).
76. Snapka, R. M., Kwok, K., Bernard, J. A., Harling, O. and Varshavsky, A. Post-separation detection of nucleic acids and proteins by neutron activation. **Proc. Natl. Acad. Sci. USA** 83:9320-9324 (1986).
82. Varshavsky, A. An electrophoretic assay for DNA-binding proteins. **Meth. Enzymol.** 151:551-565 (1987).
83. Bartel, B. and Varshavsky, A. Hypersensitivity to heavy water: a new conditional phenotype. **Cell** 52:935-941 (1988). (*An alternative to temperature-sensitive conditional mutants.*)
86. Solomon, M. J., Larsen, P. L. and Varshavsky, A. Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. **Cell** 53:937-947 (1988). (*A method for localizing specific DNA-bound proteins to specific DNA sequences through *in vivo* crosslinking, DNA fragmentation, and immunoprecipitation. This technique is known as the CHIP (chromatin immunoprecipitation) assay.*)
112. Dohmen, R. J., Wu, P. P. and Varshavsky, A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. **Science** 263:1273-1276 (1994).
113. Johnsson, N. and Varshavsky, A. Ubiquitin-assisted dissection of protein transport across membranes. **EMBO J.** 13:2686-2698 (1994).
115. Johnsson, N. and Varshavsky, A. Split ubiquitin as a sensor of protein interactions *in vivo*. **Proc. Natl. Acad. Sci. USA** 91:10340-10344 (1994).
123. Lévy, F., Johnsson, N., Rümenapf, T. and Varshavsky, A. Using ubiquitin to follow the metabolic fate of a protein. **Proc. Natl. Acad. Sci. USA** 93:4907-4912 (1996). (*A method for producing equimolar amounts of a reference and a test protein *in vivo*.*)
127. Johnson, N. and Varshavsky, A. Split ubiquitin: a sensor of protein interactions *in vivo*. In: **The Yeast Two-Hybrid System** (P. L. Bartel and S. Fields, eds.), pp. 316-332, Oxford University Press, N. Y. (1997).
137. Dünnwald, M., Varshavsky, A. and Johnsson, N. Detection of transient *in vivo* interactions between substrate and transporter during protein translocation into the endoplasmic reticulum. **Mol. Biol. Cell** 10:329-344 (1999). (*A method, based on the split-ubiquitin sensor, for detection of transient protein-protein interactions.*)
147. Varshavsky, A. The ubiquitin fusion technique and its descendants. **Meth. Enzymology** (in press).
148. Turner, G. C. and Varshavsky, A. Detecting and measuring cotranslational protein degradation *in vivo* (submitted for publication).

## The Ubiquitin System and Intracellular Proteolysis

40. Levinger, L. and Varshavsky, A. Separation of nucleosomes containing and lacking ubiquitin-H2A semihistone. **Proc. Natl. Acad. Sci. USA** 77:3244-3248 (1980).
48. Levinger, L. and Varshavsky, A. Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes in the *Drosophila* genome. **Cell** 28:375-386 (1982).

53. Swerdlow, P. and Varshavsky, A. Affinity of HMG17 for a nucleosome is not influenced by the presence of ubiquitin-H2A semihistone but depends on DNA fragment size. **Nucl. Acids. Res.** 11:387-401 (1983).
61. Finley, D., Ciechanover, A. and Varshavsky, A. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. **Cell** 37:43-55 (1984).
62. Ciechanover, A., Finley, D. and Varshavsky, A. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. **Cell** 37:57-66 (1984).
66. Özkaynak, E., Finley, D. and Varshavsky, A. The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. **Nature** 312:663-666 (1984).
70. Finley, D. and Varshavsky, A. The ubiquitin system: functions and mechanisms. **Trends Biochem. Sci.** 10:343-346 (1985).
75. Bachmair, A., Finley, D. and Varshavsky, A. *In vivo* half-life of a protein is a function of its N-terminal residue. **Science** 234:179-186 (1986).
77. Özkaynak, E., Finley, D., Solomon, M. J. and Varshavsky, A. The yeast ubiquitin genes: a family of natural gene fusions. **EMBO J.** 6:1429-1440 (1987).
78. Finley, D., Özkaynak, E. and Varshavsky, A. The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation and other stresses. **Cell** 48:1035-1046 (1987).
79. Jentsch, S., McGrath, J. P. and Varshavsky, A. The yeast DNA repair gene *RAD6* encodes a ubiquitin-conjugating enzyme. **Nature** 329:131-134 (1987).
84. Finley, D., Özkaynak, E., Jentsch, S., McGrath, J. P., Bartel, B., Pazin, M., Snapka, R. M. and Varshavsky, A. Molecular genetics of the ubiquitin system. In: **Ubiquitin** (M. Rechsteiner, ed.), pp. 39-75, Plenum Press, N. Y.(1988).
85. Varshavsky, A., Bachmair, A., Finley, D., Wünnig, I. and Gonda, D. The N-end rule of selective protein turnover: mechanistic aspects and functional implications. In: **Ubiquitin** (M. Rechsteiner, ed.), pp. 287-324, Plenum Press, N. Y. (1988).
87. Goebel, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A. and Byers, B. The yeast cell cycle gene *CDC34* encodes a ubiquitin-conjugating enzyme. **Science** 241:1331-1335 (1988).
88. Bachmair, A. and Varshavsky, A. The degradation signal in a short-lived protein. **Cell** 56:1019-1032 (1989).
89. Chau, V., Tobias, J. W., Bachmair, A., Mariott, D., Ecker, D., Gonda, D. K., and Varshavsky, A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. **Science** 243:1576-1583 (1989).
90. Finley, D., Bartel, B. and Varshavsky, A. The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. **Nature** 338:394-401 (1989).
92. Gonda, D. K., Bachmair, A., Wünnig, I., Tobias, J. W., Lane, W. S. and Varshavsky, A. Universality and structure of the N-end rule. **J. Biol. Chem.** 264:16700-16712 (1989).
94. Balzi, E., Choder, M., Chen, W., Varshavsky, A. and Goffeau, A. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. **J. Biol. Chem.** 265:7464-7471 (1990).
95. Hochstrasser, M. and Varshavsky, A. *In vivo* degradation of a transcriptional regulator: the yeast  $\alpha 2$  repressor. **Cell** 61:697-708 (1990).

96. Johnson, E. S., Gonda, D. K. and Varshavsky, A. *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. **Nature** 346:287-291 (1990).
97. Bartel, B., Wünnig, I. and Varshavsky, A. The recognition component of the N-end rule pathway. **EMBO J.** 9:3179-3189 (1990).
98. Baker, R. T. and Varshavsky, A. Inhibition of the N-end rule pathway in living cells. **Proc. Natl. Acad. Sci. USA** 88:1090-1094 (1991).
99. Varshavsky, A. Naming a targeting signal. **Cell** 64:13-15 (1991).
100. Hochstrasser, M., Ellison, M. J., Chau, V. and Varshavsky, A. The short-lived Mat $\alpha$ 2 transcriptional regulator is ubiquitinated *in vivo*. **Proc. Natl. Acad. Sci. USA** 88:4606-4610 (1991).
101. Tobias, J. W. and Varshavsky, A. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *S. cerevisiae*. **J. Biol. Chem.** 266:12021-12028 (1991).
102. Dohmen, R. J., Madura, K., Bartel, B. and Varshavsky, A. The N-end rule is mediated by the Ubc2 (Rad6) ubiquitin-conjugating enzyme. **Proc. Natl. Acad. Sci. USA** 88:7351-7355 (1991).
103. McGrath, J. P., Jentsch, S. and Varshavsky, A. *UBA1*: an essential yeast gene encoding ubiquitin-activating enzyme. **EMBO J.** 10:227-237 (1991).
104. Tobias, J. W., Shrader, T. E., Rocap, G. and Varshavsky, A. The N-end rule in bacteria. **Science** 254:1374-1377 (1991).
105. Johnson, E. S., Bartel, B., Seufert, W. and Varshavsky, A. Ubiquitin as a degradation signal. **EMBO J.** 11:497-505 (1992).
106. Ota, I. and Varshavsky, A. A gene encoding a putative tyrosine phosphatase suppresses lethality of an N-end rule-dependent mutant. **Proc. Natl. Acad. Sci. USA** 89:2355-2359 (1992).
107. Baker, R. T., Tobias, J. W. and Varshavsky, A. Ubiquitin-specific proteases of *S. cerevisiae*: cloning of *UBP2* and *UBP3*, and functional analysis of the *UBP* gene family. **J. Biol. Chem.** 267:23363-23375 (1992).
108. Varshavsky, A. The N-end rule. **Cell** 69:725-735 (1992).
109. Shrader, T. E., Tobias, J. W. and Varshavsky, A. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. **J. Bact.** 175:4364-4374 (1993).
110. Madura, K., Dohmen, R. J. and Varshavsky, A. N-recognin/Ubc2 interactions in the N-end rule pathway. **J. Biol. Chem.** 268:12046-12054 (1993).
111. Ota, I. M. and Varshavsky, A. A yeast protein similar to bacterial two-component regulators. **Science** 262:566-569 (1993).
114. Madura, K. and Varshavsky, A. Degradation of G $\alpha$  by the N-end rule pathway. **Science** 265:1454-1458 (1994).
116. Johnston, J. A., Johnson, E. S., Waller, P. and Varshavsky, A. Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. **J. Biol. Chem.** 270:8172-8178 (1995).
118. Baker, R. T. and Varshavsky, A. N-terminal amidase: a new enzyme and component of a targeting complex in the N-end rule pathway. **J. Biol. Chem.** 270:12065-12074 (1995).

119. Varshavsky, A. The world of ubiquitin. **Engineering & Science** 58:26-36 (1995).
120. Johnson, E. S., Ma, P. C. M., Ota, I. M. and Varshavsky, A. A proteolytic pathway that recognizes ubiquitin as a degradation signal. **J. Biol. Chem.** 270:17442-17456 (1995).
121. Dohmen, R. J., Stappen, R., McGrath, J. P., Forrová, H., Kolarov, J., Goffeau, A. and Varshavsky, A. An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. **J. Biol. Chem.** 270:18099-18109 (1995).
122. Varshavsky, A. The N-end rule. **Cold Spring Harbor Symp. Quant Biol.** 60:461-478 (1996).
124. Ghislain, M., Dohmen, R. J., Lévy, F., and Varshavsky, A. Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. **EMBO J.** 15:4884-4899 (1996).
125. Varshavsky, A. The N-end rule: functions, mysteries, uses. **Proc. Natl. Acad. Sci. USA** 93:12142-12149 (1996).
126. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N., Copeland, N. G. and Varshavsky, A. A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. **J. Biol. Chem.** 271:28521-28532 (1996).
128. Varshavsky, A. The N-end rule pathway of protein degradation. **Genes Cells** 2:13-29 (1997).
129. Varshavsky, A. The ubiquitin system. **Trends Biochem. Sci.** 22:383-387 (1997).
130. Varshavsky, A., Byrd, C., Davydov, I. V., Dohmen, R. J., Ghislain, M., Gonzalez, M., Grigoryev, S., Johnson, E. S., Johnsson, N., Johnston, J. A., Kwon, Y. T., Lévy, F., Lomovskaya, O., Madura, K., Rümenapf, T., Shrader, T. E., Suzuki, T., Turner, G. and Webster, A. The N-end rule pathway. In **Ubiquitin and the Biology of the Cell** (D. Finley and J.-M. Peters, eds), Plenum Press, NY, pp. 232-278 (1998).
131. Byrd, C. Turner, G. and Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. **EMBO J.** 17:269-277 (1998).
133. Davydov, I. V., Patra, D. and Varshavsky, A. The N-end rule pathway in *Xenopus* egg extracts. **Arch. Biochem. Biophys.** 357:317-325 (1998).
134. Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A. and Varshavsky, A. The mouse and human genes encoding the recognition component of the N-end rule pathway. **Proc. Natl. Acad. Sci. USA** 95:7898-7903 (1998).
135. Ramos, P. C., Höckendorff, J., Johnson, E. S., Varshavsky, A. and Dohmen, R. J. Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. **Cell** 92:489-499 (1998).
136. Lévy, F., Johnston, J. A. and Varshavsky, A. Analysis of a conditional degradation signal in yeast and mammalian cells. **Eur. J. Biochem.** 259:244-252 (1999).
138. Kwon, Y. T., Kashina, A. S. and Varshavsky, A. Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. **Mol. Cell. Biol.** 19:182-193 (1999).
140. Kwon, Y. T. Lévy, F. and Varshavsky, A. Bivalent inhibitor of the N-end rule pathway. **J. Biol. Chem.** 274:18135-18139 (1999).

141. Xie, Y. and Varshavsky, A. The N-end rule pathway is required for import of histidine in yeast lacking the kinesin-like protein Cin8p. **Curr. Genet.** 36:113-123 (1999).
142. Xie, Y. and Varshavsky, A. The E2-E3 interactions in the N-end rule pathway: The RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain. **EMBO J.** 18:6832-6844 (1999).
143. Suzuki, T. and Varshavsky, A. Degradation signals in the lysine-asparagine sequence space. **EMBO J.** 18:6017-6026 (1999).
144. Xie, Y. and Varshavsky, A. Physical association of ubiquitin ligases and the 26S proteasome. **Proc. Natl. Acad. Sci. USA** 97:2497-2502 (2000).
145. Kwon, Y. T., Balogh, S. A., Davydov, I. V., Kashina, A. S., Yoon, J. K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V. H. and Varshavsky, A. Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. **Mol. Cell. Biol.** 20:4135-4148 (2000).
146. Turner, G. C., Du, F. and Varshavsky, A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. **Nature** (in press).
149. Davydov, I. V. and Varshavsky, A. RGS4 is arginylated and degraded by the N-end rule pathway *in vitro*. **J. Biol. Chem.** (in press).

### **Multitarget Compounds**

117. Varshavsky, A. Codominance and toxins: a path to drugs of nearly unlimited selectivity. **Proc. Natl. Acad. Sci. USA** 92:3663-3667 (1995).
132. Varshavsky, A. Codominant interference, antieffectors, and multitarget drugs. **Proc. Natl. Acad. Sci. USA** 95:2094-2099 (1998).

## **U. S. Patents**

1. Varshavsky, A. *Gene amplification assay for detecting tumor promoters.* Issued Apr. 10, 1984.
2. Snapka, R. M., Kwok, K. S., Bernard, J. A., Harling, O. R. and Varshavsky, A. *Indirect labeling method for post-separation detection of chemical compounds.* Issued Dec. 10, 1991.
3. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for generating desired amino-terminal residues in proteins.* Issued Mar. 3, 1992.
4. Varshavsky, A., Johnson, E. S., Gonda, D. K. and Hochstrasser, M. *Methods for trans-destabilization of specific proteins in vivo and DNA molecules useful thereof.* Issued Jun. 16, 1992.
5. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for producing proteins and polypeptides using ubiquitin fusions.* Issued Jul. 21, 1992.
6. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for in vitro cleavage of ubiquitin fusion proteins.* Issued Mar. 23, 1993.
7. Baker, R. T., Tobias, J. W. and Varshavsky, A. *Nucleic acid encoding ubiquitin-specific proteases.* Issued May 18, 1993.
8. Tobias, J. W. and Varshavsky, A. *Ubiquitin-specific protease.* Issued Feb. 21, 1995.
9. Baker, R. T., Tobias, J. T. and Varshavsky, A. *Ubiquitin-specific proteases.* Issued Feb. 27, 1996.
10. Johnsson, N. and Varshavsky, A. *Split-ubiquitin protein sensor.* Issued Apr. 2, 1996.
11. Wu, P., Dohmen, R. J., Johnston, J. and Varshavsky, A. *Heat-inducible N-degron module.* Issued Jul. 23, 1996.
12. Johnston, J. and Varshavsky, A. *Inhibiting degradation of a degron-bearing protein.* Issued Jun. 9, 1998.
13. Baker, R. T., Gonda, D. K. and Varshavsky, A. *Inhibition of protein degradation in living cells with peptides.* Issued June 16, 1998.
14. Kwon, Y. T. and Varshavsky, A. *Nucleic acid encoding mammalian Ubr1.* Issued Jan. 19, 1999.

# **Conditional toxin against yeast cells expressing HIV protease**

Ilia V. Davydov, and Alexander Varshavsky\*

Division of Biology,  
California Institute of Technology,  
Pasadena, California 91125

*\*Corresponding author:* Alexander Varshavsky, Division of Biology, 147-75,  
California Institute of Technology,  
1200 East California Boulevard, Pasadena, CA 91125

*Telephone:* 626-395-3785; *Fax:* 626-440-9821

*Email:* avarsh@cco.caltech.edu

*Running title:* Toxin activated by HIV PR

## ABSTRACT

A cleavage-activated conditional toxin is an engineered protein fusion, whose toxicity is enhanced as a result of proteolytic cleavage by a viral protease (PR). A conditional toxin of this type contains (a) a cytotoxic domain; (b) an inactivating domain; and (c) a PR cleavage site located between the cytotoxic and the inactivating domains. The presence of the inactivating domain greatly reduces the toxicity of the cytotoxic domain. Cleavage of the fusion by PR leads to the release of the toxic domain, resulting in its higher toxicity. We constructed a cleavage-mediated conditional toxin with enhanced toxicity against yeast cells expressing PR of the human immunodeficiency virus (HIV PR). In our design the A fragment of diphtheria toxin (DTA) serves as a cytotoxic domain and the yeast transcriptional repressor Cup9 serves as an inactivating domain. Cup9 is a metabolically unstable protein with predicted nuclear localization. The cleavage-specific toxic effect of the Cup9-DTA-based conditional toxin that we observed may have two underlying contributing mechanisms. First, cleavage may lead to metabolic stabilization and, as a result, intracellular accumulation of DTA. Second, cleavage may interfere with nuclear localization of DTA leading to its greater toxicity. Our system represents a fast experimental approach to construct and test conditional toxins directed against HIV PR-containing cells.

## INTRODUCTION

The human immunodeficiency virus (HIV) protease (PR) is essential at the later stages of the replication cycle of the virus. It is responsible for proteolytic cleavage of Gag and Gag-Pol polyprotein precursors at specific cleavage sites [Dougherty, 1993 #270]. Treatment of HIV infection with PR inhibitors in combination with other antiretroviral drugs can produce sustained suppression of HIV replication and reduction of viral loads. However, the virus continues to exist latently in some infected cells, and this represents a major hurdle to its complete eradication [Wong, 1997 #267; Finzi, 1997 #268]. In the long run, the emergence of drug-resistant strains of HIV may create another therapeutic problem. Therefore, some research groups focused on development of an alternative anti-HIV approach that would ensure death of HIV-infected cells thus eliminating endogenous viral pools. Thus, it was reported that fusions of toxic proteins and human CD4 molecule display selective toxicity toward cells expressing the HIV envelope glycoprotein and therefore represent a potential therapeutic tool to kill HIV-infected cells [Aljaufy, 1994 #269; Aullo, 1992 #271; Till, 1988 #272; Chaudhary, 1988 #273].

We previously proposed an idea of a new class of conditional toxins (termed *sitoxins* for signal-regulated, cleavage-mediated *toxins*) that contain a toxic domain linked to a signal that inactivates it [Varshavsky, 1996 #89]. In the presence of HIV PR the toxic domain would be released from the inactivating signal and becomes significantly more toxic. We proposed that a degradation signal or a nuclear localization domain could serve as an inactivating signal of a sitoxin. Recently it was reported that an

engeneered version of the apoptosis-promoting caspase-3 can be activated by the HIV PR, leading to apoptosis of the HIV-infected cells [Vocero-Akbani, 1999 #265].

In the present work we describe the construction of a conditional toxin based on the fusion between the yeast transcriptional repressor Cup9 and the A fragment of diphtheria toxin (DTA) linked by the polypeptide sequence containing an HIV PR cleavage site. This conditional toxin dispalys increased toxicity against yeast cells expressing HIV PR. Our system may serve as a model for construction and quick testing of HIV PR-directed conditional toxins.

## EXPERIMENTAL PROCEDURES

*Plamids and the yeast strain* - p416Gal1 and p414Met25 yeast expression vectors [Mumberg, 1994 #279] were a generous gift of M. Funk (Institute of Molecular Biology, Marburg, Germany). pSE979 plasmid containing DTA ORF [Bellen, 1992 #286] was a kind gift of S.J. Elledge (Howard Hughes Medical Institute). The construct NefPR [Arrigo, 1995 #280] was a kind gift of S.J. Arrigo (Medical University of South Carolina).

The plasmid p416Gal1-Cup9-LHIV-DHFR-myc drives the expression of Cup9-LHIV-DHFR-myc fusion protein in yeast. The Cup9-LHIV-DHFR-myc ORF was cloned between the *Xba*I and *Hind*III sites of the vector p416Gal1. The full-length Cup9 ORF flanked with restriction sites *Xba*I at the 5'-end and *Bam*HI at the 3'-end was generated by PCR from yeast genomic DNA. The LHIV linker encoding HIV PR cleavage site VSFNF/PQITL and flanked by restriction sites *Bam*HI at the 5'-end and *Bgl*II at the 3'-end was generated by inserting the oligonucleotide duplex 5'-

AGGATCCGTATCCTTAACCTCCCTCAGATCACTCTGGCAGATCTGGTAC-3'

and

5'-CAGATCTGCCAAGAGTGATCTGAGGGAAGTTAAAGGATACGGATCCTG-TAC-3'. The full-length mouse DHFR sequence was flanked with *Bgl*II site at the 5'-end and *Kpn*I site at the 3'-end. The linker encoding myc epitope with terminator codon at it's end was flanked by *Kpn*I site at it's 5'-end and *Hind*III site at the 3'-end. The complete details of the construction procedure are available upon request.

p416Gal1-Cup9(82-306)-LHIV-DHFR-myc was constructed by replacing the full-length Cup9 ORF in p416Gal1-Cup9-LHIV-DHFR-myc between *Xba*I and *Bam*HI sites with the truncated Cup9(82-306) ORF generated by PCR. p416Gal1-Cup9(82-306)-LHIV-DTA was produced by replacing the DHFR-myc-encoding fragment between the restriction sites *Bgl*II and *Xho*I in p416Gal1-Cup9(82-306)-LHIV-DHFR-myc (*Xho*I site is present downstream from *Hind*III site in p416Gal1 vector) with DTA-encoding fragment. In p416Gal1-Cup9(82-306)-LHIVmut-DTA the sequence 5'-GTATCCTTAACCTCCCTCAGATCACTCTT-3' encoding HIV PR cleavage site was changed to 5'-GTATCCTTAACGTTCCCTCAGATCACTCTT-3'.

Construction of the plasmid used for expressing HIV PR in yeast involved the following steps. The NefPR plasmid for HIV PR expression in mammalian cells was digested with *Spe*I and two copies of the following oligonucleotide duplex encoding flag epitope were inserted: 5'-CTAGTGGCGACTACAAGGACGACGATGACAAGT-3' and 5'-CTAGACTTGTCACTCGTCGTCCCTGTAGTCGCCA-3'. The resulting construct pNef(flag)<sub>2</sub>PR was used to asses the efficiency of self-cleavage by HIV PR. To subclone the Nef(flag)<sub>2</sub>PR ORF into a yeast expression vector we introduced a *Eco*RI site upstream

from the 5'-end of the ORF, excised it with *Eco*RI enzyme (the second *Eco*RI site is present at the 3'-end of the ORF) and subcloned it into the p414Met25 plasmid for yeast expression.

The yeast strain used in current work was JD47-13C [Madura, 1993 #61].

*Western Blot, immunoprecipitation and pulse-chase assays* - were done essentially as previously described [Ghislain, 1996 #126]. Anti-myc monoclonal antibodies 9E10 were purchased from Berkeley Antibody company (CA).

*Yeast viability assays* - JD47-13C yeast were transformed simultaneously with two plasmids: p416Gal1-Cup9(82-306)-LHIV-DTA, p416Gal1-Cup9(82-306)-LHIVmut- DTA or p416Gal1 as a first plasmid; and p414Met25-Nef(flag)<sub>2</sub>PR or p414Met25 as a second plasmid. The transformation procedure involved yeast competent cells stored at - 80°C [Dohmen, 1991 #287]. Primary transformants were grown on plates containing 0.67% yeast nitrogen base without amino acids (Difco) supplemented with auxotrophic amino acids His, Leu and Lys, 2% glucose and 1 mM Met for repression of the Met25 promoter (growth medium) with 2% agar. Colonies were picked and grown 24 hours in 2 ml of the liquid growth medium. Next day the cultures were diluted to OD<sub>600</sub>~0.2 in the growth medium and grown until OD<sub>600</sub>~0.5-1.5. Aliquots of about 0.65 optic units (1 ml of the culture with OD<sub>600</sub>~0.65) were taken from each culture, washed twice with the induction medium (0.67% yeast nitrogen base supplemented with His, Leu and Lys and 2% galactose) and resuspended in 430 µl of the induction medium. The cultures were left on the bench for 45 min and spread on Petri dishes containing the induction medium with 2% agar using a bacteriological loop. Yeast growth on the plates was monitored for 5 days.

## RESULTS AND DISCUSSION

The general idea of a sitoxin is illustrated on Figure 1. The crucial property of a sitoxin is that its toxicity is greatly enhanced as a result of proteolytic cleavage by HIV PR. In general, a conditional toxin of this type is a protein fusion that contains (1) a cytotoxic domain; (2) an inactivating domain (see below); and (3) a domain located between the two aforementioned domains and bearing the recognition and cleavage sites of HIV PR. The presence of the inactivating domain greatly reduces the toxicity of the cytotoxic domain, ideally making it non-toxic as a part of the fusion. In the presence of HIV PR the fusion would be cleaved and the toxic domain released from the fusion, leading to its greatly enhanced toxicity.

One may envision different possible mechanisms of inhibiting the toxicity of the cytotoxic domain by the inactivating domain. In the first possible mechanism, the inactivation domain may contain a degradation signal, making the whole fusion short-lived intracellularly, and therefore relatively non-toxic. The HIV PR would cleave the fusion resulting in metabolic stabilization of the toxin and therefore its intracellular accumulation and higher toxicity. In the second possible mechanism, the inactivation domain may contain a translocation signal, such as NLS. If the action of the toxin is cytosol-specific and the NLS-toxin fusion is rapidly translocated into the nucleus, then the toxic domain would be unable to exert its toxic effect. Cleavage by the HIV PR would result in the physical uncoupling of NLS and the toxic domain. Having lost the NLS, the toxic domain would remain in the cytosol and become significantly more toxic.

Some other mechanisms of action of the inactivation domain of a conditional toxin, including sterical masking of the active site of a toxin, are possible as well.

We decided to construct a version of conditional toxin based on a fusion between the yeast homeodomain protein Cup9 as inactivation domain and DTA as toxic domain. Cup9 is a metabolically unstable protein with a half-life 5-10 min, which is degraded by the Ubr1-dependent degradation pathway (also referred to as “the N-end rule pathway”) in yeast [Byrd, 1998 #274]. Cup9 also acts as a transcriptional repressor of the *PTR2* gene, encoding a peptide transporter [Byrd, 1998 #274]. Cup9 is predicted to be a nuclear protein functionally. It also has three positively charged clasters of amino acids at positions 156 to 160, 217 to 220, and 286-289; similar clasters have been identified to be important in the localization of nuclear proteins [Knight, 1994 #95]. We reasoned that a Cup9-based inactivation domain would combine the properties of both types of the inactivation signals described above, inhibiting the action of the toxic domain by both metabolic destabilization and translocation into the nucleus.

DTA is the enzymatically active fragment of diphtheria toxin. It inhibits cellular protein synthesis in eukaryotes by inactivating elongation factor 2 through ADP-ribosylation [Moss, 1990 #277]. Therefore, it is expected to be significantly less toxic when directed for localization into the nucleus. Yeast were shown to be susceptible to the DTA toxicity [Chen, 1985 #276], making them a convenient model to demonstrate the sitoxin principle.

The HIV PR cleavage site engineered between the Cup9 and DTA domains has the sequence VSFNF/PQITL, which represents the cleavage site at the NH<sub>2</sub>-terminus of

the HIV PR in the Gag-Pol polyprotein precursor, which is cleaved most efficiently [Krausslich, 1989 #275].

Figure 2 shows schematically the recombinant genes used in the present study. Initially we constructed the protein fusion named Cup9-LHIV-DHFR-myc. In this name Cup9 represents the full-length Cup9 protein; LHIV denotes the 19 amino acid (aa) linker containing the VSFNF/PQITL cleavage site of the HIV PR; DHFR stands for the full-length mouse DHFR aa sequence; and myc is the epitope for immunodetection. We intended to use this fusion as a reporter in preliminary experiments by coexpressing it together with the HIV PR and monitoring the efficiency of its cleavage in Western blots and immunoprecipitations with anti-myc antibodies. However, the Cup9-LHIV-DHFR-myc appeared to be toxic when overexpressed in yeast apparently because of the presence of the full-length Cup9 moiety, since Cup9-myc was also significantly toxic upon overexpression (Fig. 3; and C. Byrd, I. V. D., and A. V., unpublished data). We decided to modify the Cup9 sequence to reduce its toxicity. Preliminary evidence indicated that the degradation signal of Cup9 was located in the C-terminal portion of the protein (C. Byrd, I. V. D., and A. V., unpublished data). Therefore, we constructed truncated versions of the original Cup9-LHIV-DHFR-myc reporter fusion containing deletions from the N-terminus of Cup9. In this way we were attempting to obtain a less toxic reporter protein that would retain the Cup9 degradation signal. One of the thus modified fusions, named Cup9(82-306)-LHIV-DHFR-myc, had the 81 N-terminal aa of Cup9 deleted (Fig. 2). Upon overexpression in yeast under the control of Gal1 promoter from the centromeric p416Gal1 vector [Mumberg, 1994 #279], this truncated fusion was

significantly less toxic than the original Cup9-LHIV-DHFR-myc (Fig. 3). Therefore, in the subsequent experiments we used the Cup9(82-306)-LHIV-DHFR-myc reporter.

As a source of HIV PR we employed the chimeric protein NefPR [Arrigo, 1995 #280]. NefPR contains the complete HIV PR sequence flanked on both sides with the sites for self-cleavage and embedded into the heterogenous aa sequence of the HIV Nef protein. As a part of this fusion the HIV PR was shown to be efficiently expressed and self-cleaved in mammalian cells [Arrigo, 1995 #280; Arrigo, 1995 #278]. We added the double flag epitope to the Nef portion of the fusion and confirmed that the Nef domain was released as a result of the self-cleavage of the HIV PR (data not shown). Then we placed the Nef(flag)<sub>2</sub>PR-encoding ORF under the control of the Met25 promoter into the yeast vector p414Met25 [Mumberg, 1994 #279] (Fig. 2) and used the resulting construct for the next experimental steps.

The Cup9(82-306)-LHIV-DHFR-myc reporter was expressed alone or together with Nef(flag)<sub>2</sub>PR in yeast cells and its cleavage was studied in Western blots with anti-myc antibodies (Fig. 4A). In the absence of HIV PR, Cup9(82-306)-LHIV-DHFR-myc was migrating as a single band in the area of the gel consistent with its deduced molecular weight of 49 kD. In the presence of HIV PR this band decreased in intensity and the new very prominent faster migrating band corresponding to the DHFR-myc cleavage product appeared (Fig. 4A). In another experiment we pulse-labeled the yeast cells for 1 hour, lysed them and immunoprecipitated the proteins with the anti-myc antibodies (Fig. 4B). This experiment confirmed the results of the Western blot. During 1 hour of labeling in the presence, but not in the absence of HIV PR the full-length Cup9(82-306)-LHIV-DHFR-myc reporter was cleaved and the cleavage product

accumulated to significant amounts (Fig. 4B). We concluded that the reporter Cup9(82-306)-LHIV-DHFR-myc was cleaved efficiently by HIV PR in our system.

We were somewhat surprised that the intensity of the band of the cleavage product on the Western blot and immunoprecipitation appeared much stronger than the intensity of the band of the original full-length fusion. There are two possible explanations for this observation. First, the full-length Cup9-containing reporter might be tightly bound to DNA in the nucleus, and our extraction procedure of vortexing with glass beads in buffer containing 1% Triton X-100 and 0.15 M NaCl might not be suitable for its complete recovery. Second, because the full-length reporter fusion is expected to be an unstable protein due to the presence of a degradation signal inside Cup9, its low level may be due to its relatively fast degradation; in contrast, since DHFR-myc cleavage product is metabolically stable, it may accumulate. To address the first possibility, we employed an alternative extraction procedure of preparing yeast spheroplasts by treatment with lyticase and subsequently lysing them in the SDS-loading buffer. This very vigorous extraction protocol produced the same result on the Western blot as our original method (data not shown). Although the possibility of incomplete extraction of the full-length fusion can not be completely excluded, this experiment speaks against it. To address the second possibility, we performed pulse-chase analysis (Fig. 4C). During the course of the 30 min chase in the presence of HIV PR the DHFR-myc product accumulated significantly. Without HIV PR the reporter fusion appeared only slightly unstable in this pulse-chase assay (about 80% left after 30 min of chase). Therefore, the increased intensity of the band of the DHFR-myc truncated product compared to the band of the full-length reporter can not be entirely due to their differential metabolic stabilities.

Currently we do not have a full explanation of this phenomenon. In any case, we concluded that the reporter protein Cup9(82-306)-LHIV-DHFR-myc was efficiently cleaved by HIV PR. Therefore, we proceeded to the construction of a conditional toxin based on this reporter fusion.

To construct a conditional toxin, we replaced the DHFR-myc portion of the reporter fusion Cup9(82-306)-LHIV-DHFR-myc with DTA. The resulting construct was called Cup9(82-306)-LHIV-DHFR-myc (or simply Tox) and was used as the working conditional toxin. In addition, we constructed a mutant version of this conditional toxin by replacing the HIV PR cleavage site VSFNF/PQITL with VSFNV/PQITL. The F to V point mutation at the cleavage junction was previously reported to inhibit the HIV PR cleavage dramatically [Loeb, 1989 #282]. We confirmed that in the context of our DHFR-containing reporter fusion the mutated cleavage site was cleaved very inefficiently (data not shown). The mutated conditional toxin bearing the F to V point mutation at the HIV PR cleavage site was named Cup9(82-306)-LHIVmut-DTA (or simply mTox). Tox and mTox were expressed in yeast under the control of Gal1 promoter from the centromeric vector p416Gal1 in the presence or in the absence of HIV PR\*. After growing the cells in glucose in the presence of methionine, the expression of Tox, mTox and HIV PR was induced and viability of the cells measured by growing the yeast on galactose-containing minimal medium without methionine on Petri plates. We observed that coexpression of Tox, but not mTox with HIV PR resulted in significant increase of toxicity (Fig. 5). The fact that the F to V substitution at the HIV PR cleavage

---

\* DTA-containing plasmids had dramatically lower efficiencies of transfecting yeast. This might be due to a spike of transient unregulated expression of the protein fusions during or immediately after the transfection procedure.

site interfered with the toxicity indicates that the toxic effect we observed was cleavage-mediated. Thus, the fusion protein Cup9(82-306)-LHIV-DTA works as a cleavage-mediated conditional toxin.

As was described above, the cleavage-specific toxic effect that we observed may have two underlying contributing mechanisms. First, cleavage may lead to metabolic stabilization and, as a result, intracellular accumulation of DTA. Second, cleavage may interfere with nuclear localization of DTA leading to its greater toxicity. Both of these possibilities are likely and may, in fact, contribute together to the observed cleavage-mediated toxicity. In addition, it can not be excluded that the Cup9 portion of the toxin interferes with DTA toxicity.

Our system represents a fast experimental approach to construct and test a conditional toxin directed against cell containing HIV PR. Recently Vocero-Akbani et al. reported that a modified apoptotic protease caspase-3 containing an HIV PR cleavage site demonstrates preferential toxicity against HIV-infected cells. The work of these authors apparently presents another attempt to construct a cleavage-mediated conditional toxin. Taken together, these attempts could eventually lead to creating a clinically useful toxic molecule to eradicate HIV-infected cells.

A modification of our system could also be useful as a genetic test for enzymatic activities of different mutant versions of HIV PR. Thus, HIV PR sequences from different HIV isolates could be PCR-amplified, inserted into the yeast expression vector, and their enzymatic activities estimated by co-expression with a conditional toxin and monitoring the yeast cell survival. In yet another possible application of this system,

it could be used to measure efficiencies of recognition and cleavage of different sites by HIV PR.

In sum, we constructed a cleavage-mediated conditional toxin and demonstrated it's preferential toxicity against HIV PR-expressing yeast.

#### ACKNOWLEDGEMENTS

We would like to thank M. Funk for the gift of p416 Gal1 and p414Met25 yeast expression vectors; S.J. Elledge for providing the DTA ORF-containing plasmid; and S.J. Arrigo for the NefPR plasmid. We are thankful to the members of A. Varshavsky's lab for critical discussions during this work and for comments on the manuscript.

## FIGURE LEGENDS

**FIGURE 1. Cleavage-mediated conditional toxin (sitoxin).** In the sitoxin idea the toxicity of a protein fusion is enhanced as a result of proteolytic cleavage by HIV PR. HIV PR cleavage removes the inactivating domain that inhibits the activity of the cytotoxic domain as a part of the fusion.

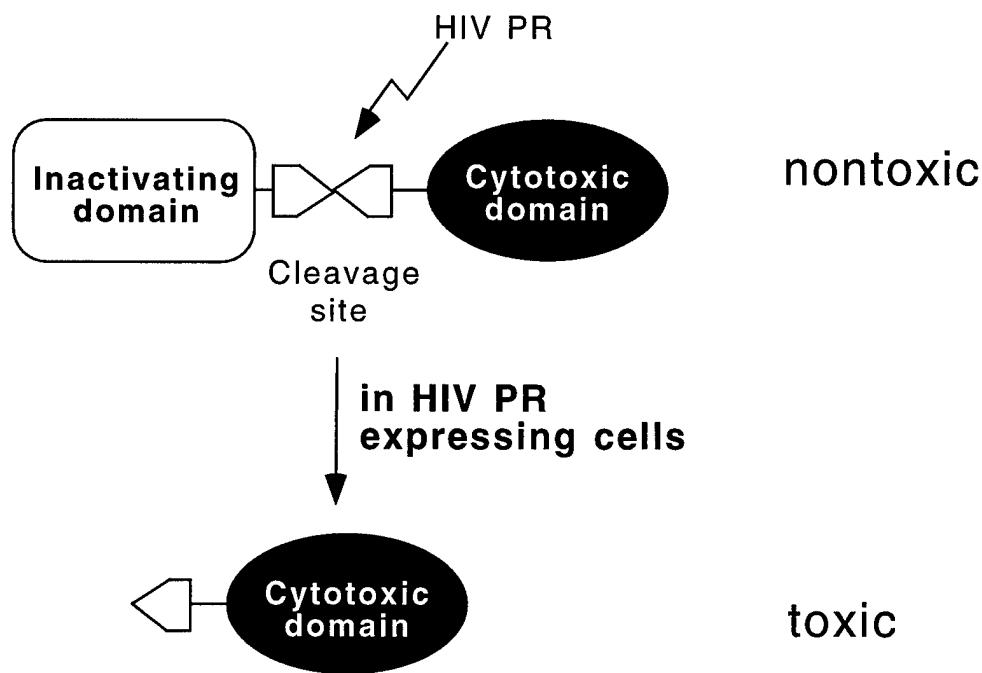
**FIGURE 2. Constructs used in the current study.** Arrows over the Nef(flag)<sub>2</sub>PR construct indicate the sites of self-cleavage of HIV PR from the fusion. Sizes of the boxes on this illustration were not intended to be exactly proportional to the sizes of the corresponding DNA fragments.

**FIGURE 3. Toxicity of Cup9-containing fusions upon overexpression in yeast.** Yeast were transformed with recombinant plasmids encoding the corresponding protein fusions under the control of Gal1 promoter or an empty vector as indicated. Yeast cultures were processed as described in Experimental Procedures and distributed over Petri plates with galactose-containing medium to induce protein expression. Picture was taken after 4 days of growth on the Petri plate.

**FIGURE 4. Cleavage by the HIV PR of the reporter fusion Cup9(82-306)-LHIV-DHFR-myc upon co-expression in yeast.** The black arrow indicates the original Cup9(82-306)-LHIV-DHFR-myc protein fusion. The white arrow indicates its cleavage product DHFR-myc. Molecular weights of the markers (in kDa) are indicated on the

right side of each panel. *a.* Western blot with anti-myc antibodies. The reporter fusion Cup9(82-306)-LHIV-DHFR-myc was expressed either alone (lane 1) or together with HIV-PR (lane 2). *b.* Immunoprecipitations with anti-myc antibodies (Berkeley Antibody Company, CA). Yeast expressing no foreign proteins (lane 1), the reporter fusion Cup9(82-306)-LHIV-DHFR-myc alone (lane 2) or together with HIV PR (lane 3) were labeled for 1 hour with a mixture of  $^{35}$ S-methionine and  $^{35}$ S-cystein. Protein extracts were prepared, immunoprecipitated with anti-myc antibodies and resolved on SDS-PAGE. *c.* Pulse-chase assays. Yeast cells were expressing Cup9(82-306)-LHIV-DHFR-myc reporter protein and HIV-PR as indicated. Yeast were pulse-labeled with a mixture of  $^{35}$ S-methionine and  $^{35}$ S-cystein for 165 sec and chased in the excess of cold methionine and cysteine and 200  $\mu$ g/ml cyclohexamide for 0, 10 and 30 min as marked. After the chases protein extracts were prepared, immunoprecipitated with anti-myc antibodies and resolved on SDS-PAGE.

**FIGURE 5. Toxicity of Tox, but not mTox, is enhanced in the presence of HIV PR.** Yeast were transformed with recombinant plasmids encoding the corresponding proteins (+) or empty vectors (-) as indicated. Yeast cultures were processed as described in Experimental Procedures and distributed over Petri plates with galactose-containing medium to induce Tox and mTox under the control of Gal1 promoter and no methionine to induce HIV PR under the control of Met25 promoter. Pictures were taken after 5 days of growth on Petri plates.





**Cup9-LHIV-DHFR-myc**



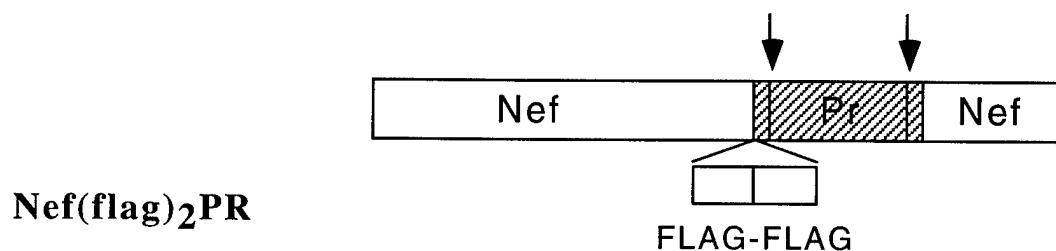
**Cup9(82-306)-LHIV-DHFR-myc**



**Cup9(82-306)-LHIV-DTA**

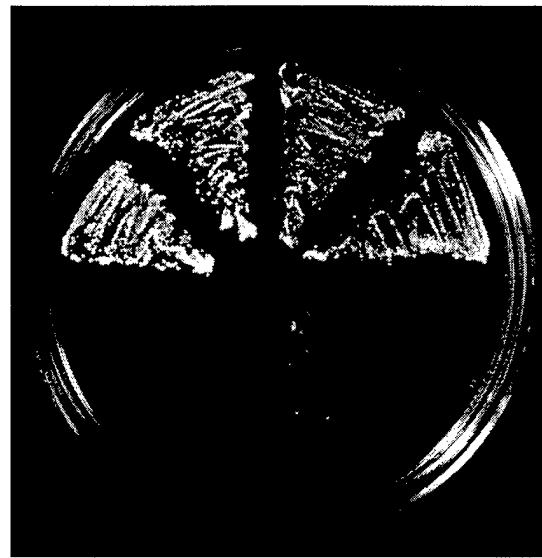


**Cup9(82-306)-LHIVmut-DTA**



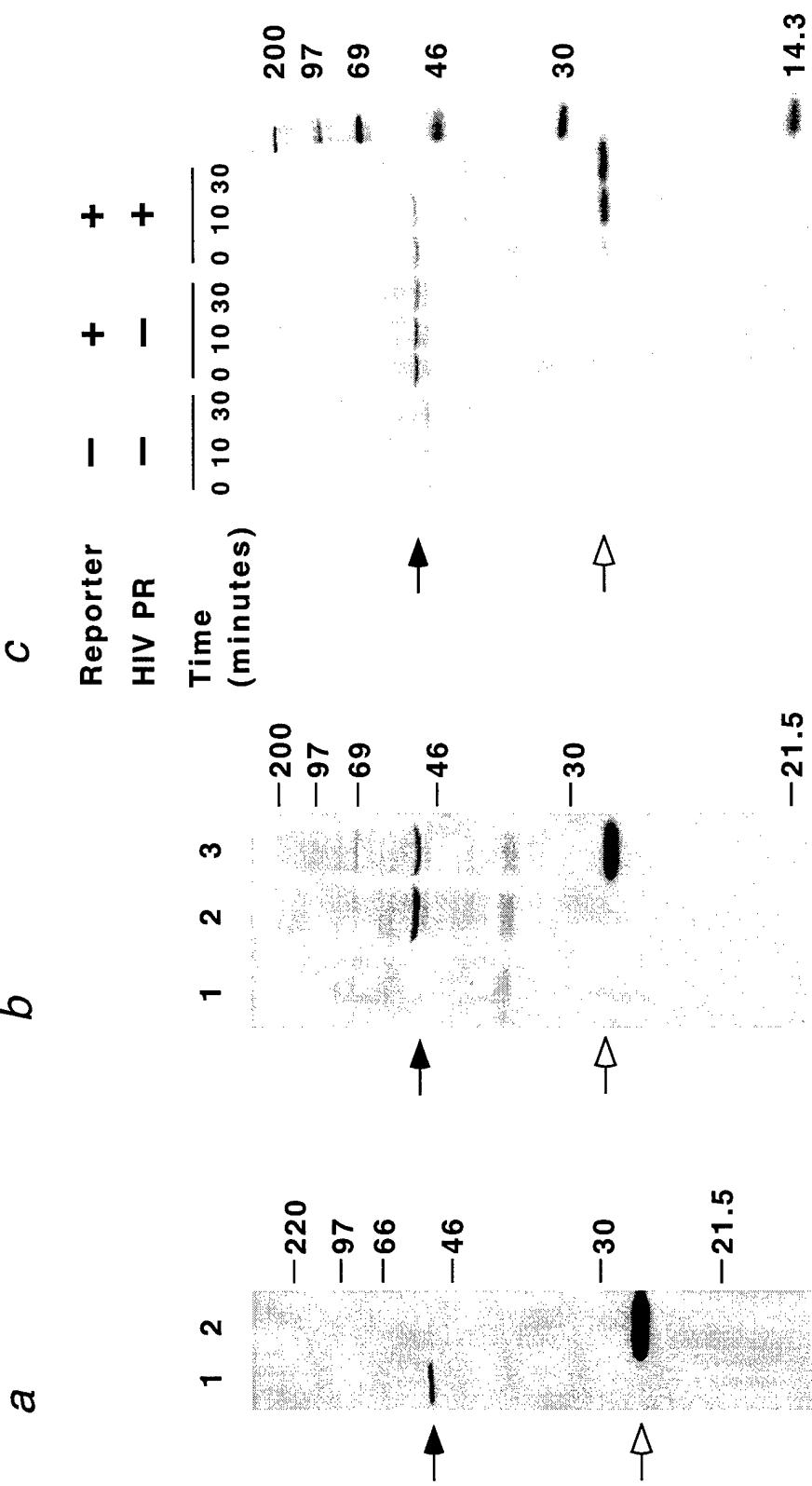
**Nef(flag)<sub>2</sub>PR**

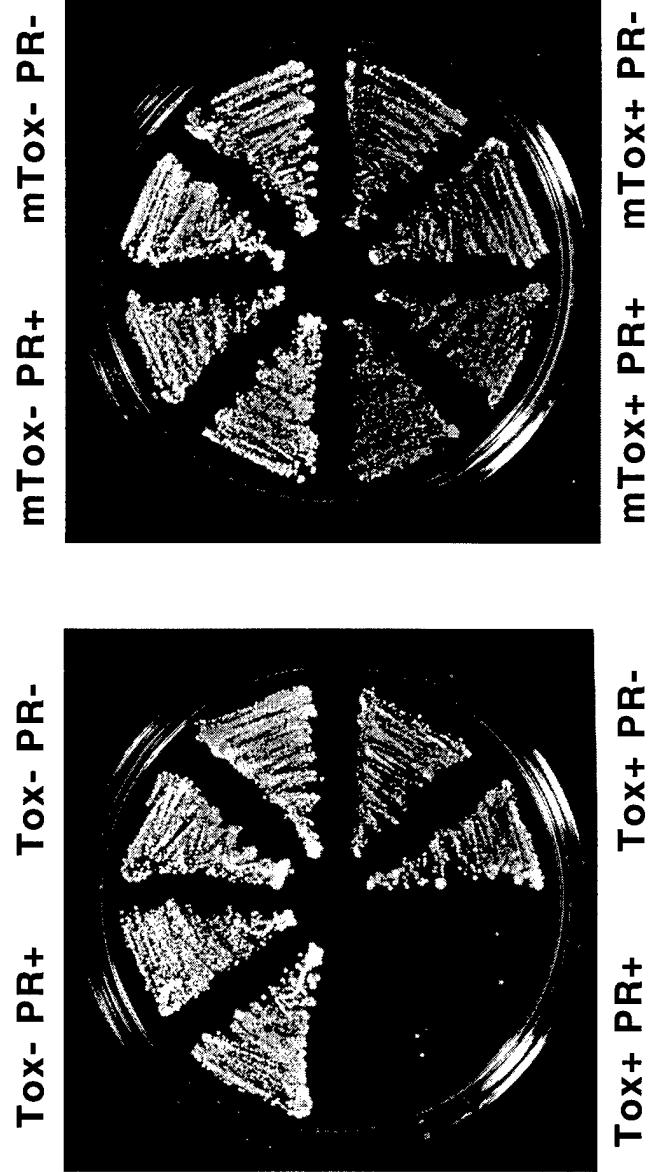
**Cup9-myc**



**vector**

**Cup9(82-306)-  
-LHIV-DHFR-myc**





*Letter to Nature*  
Revised

**Peptides accelerate their uptake by activating a ubiquitin-dependent  
proteolytic pathway**

Glenn C. Turner<sup>†</sup>, Fangyong Du<sup>†</sup>, and Alexander Varshavsky\*

Division of Biology,  
California Institute of Technology,  
Pasadena, California 91125, USA

\*Corresponding author: Alexander Varshavsky, Division of Biology, 147-75,  
Caltech, 1200 East California Boulevard, Pasadena, CA 91125, USA

<sup>†</sup>These authors contributed equally to this work.

Telephone: 626-395-3785; Fax: 626-440-9821; Email: avarsh@caltech.edu

*Keywords:* proteolysis / ubiquitin / proteasome / N-end rule / peptide import

Protein degradation by the ubiquitin (Ub) system controls the intracellular concentrations of many regulatory proteins. The degradation signals presented by these proteins are recognized by the E3 components of the Ub system<sup>1-3</sup>. Here we describe the first example of physiological regulation of a Ub-dependent pathway through allosteric modulation of its E3 activity by small compounds. Ubr1p, the E3 of the N-end rule pathway in *Saccharomyces cerevisiae*, mediates the Ub/proteasome-dependent degradation of Cup9p, a transcriptional repressor of the peptide transporter Ptr2p<sup>4</sup>. Ubr1p also targets proteins that bear destabilizing N-terminal residues<sup>5</sup>. Here we show that the degradation of Cup9p is allosterically activated by dipeptides bearing destabilizing N-terminal residues. In the resulting positive feedback circuit, imported dipeptides bind to Ubr1p and accelerate the Ubr1p-dependent degradation of Cup9p, thereby derepressing the expression of Ptr2p and increasing the cell's capacity to import peptides. These findings identify the physiological rationale for the targeting of Cup9p by Ubr1p, and suggest that small compounds may regulate other Ub-dependent pathways as well.

A protein target of the Ub system is conjugated to Ub through the action of three enzymes, E1, E2 and E3<sup>2,3,6</sup>. The resulting ubiquitylated protein bears a covalently linked multi-Ub chain, and is degraded by the 26S proteasome<sup>7</sup>. The selectivity of ubiquitylation is determined mainly by the E3, which recognizes a degradation signal (degron) carried by the target protein. The rate of degradation of specific proteins is often regulated, in response either to signals from the environment or to signals produced intracellularly, for example, during cell cycle progression. In many cases, this is achieved by modulating the exposure or the structure of a degron in a target protein. For example, the degrons of the cyclin-dependent kinase inhibitors Sic1p and p27 are activated by phosphorylation, which is timed to bring about their destruction at key transition points in the cell cycle<sup>8</sup>. In other cases, phosphorylation regulates the activity of an E3 itself. For example, the anaphase-promoting complex (APC), a multisubunit E3, is activated only at mitosis<sup>9</sup>.

One Ub-dependent proteolytic system, termed the N-end rule pathway, targets proteins carrying a degradation signal called the N-degron<sup>10</sup>. The essential determinants of an N-degron are a destabilizing N-terminal residue and a lysine residue of the substrate<sup>11</sup>. The N-end rule pathway is present in all organisms examined, from mammals and fungi to bacteria<sup>12</sup>. In *S. cerevisiae*, there are two classes of destabilizing residues, basic, or type 1 (Arg, Lys and His) and bulky hydrophobic, or type 2 (Phe, Leu, Tyr, Trp and Ile). Ubr1p, a 225K, RING-H2 finger-containing E3, directly recognizes these N-terminal residues<sup>5,13</sup>. The lysine determinant of an N-degron is the site of formation of a substrate-linked multi-Ub chain, which is produced by a complex of the E3 Ubr1p and the E2 enzyme Rad6p (Ubc2p)<sup>14</sup>. Dipeptides bearing a destabilizing N-terminal residue of either basic or hydrophobic type act as competitive inhibitors of the degradation of N-end rule substrates carrying the same type of destabilizing residue<sup>15-17</sup>. Thus Ubr1p contains two distinct N-terminal residue-binding sites that are each capable of binding either a dipeptide or a protein, but not both at the same time.

The N-end rule pathway was discovered through the use of engineered reporter proteins<sup>10</sup>. Recent work identified the first physiological function of Ubr1p in *S. cerevisiae*, by demonstrating that Ubr1p is required for peptide import<sup>18</sup>. Ubr1p regulates the uptake of peptides by controlling degradation of the 35K homeodomain protein Cup9p, a transcriptional repressor of the di- and tripeptide transporter Ptr2p<sup>4</sup>. In contrast to the canonical N-end rule substrates, Ubr1p targets Cup9p through a degron located in the C-terminal half of Cup9p (F. Navarro-Garcia, G. Turner and A. Varshavsky, unpublished data). Despite this unexpected mode of recognition, we asked whether dipeptides bearing destabilizing N-terminal residues could affect the Ubr1p-mediated degradation of Cup9p, since dipeptides are able to inhibit degradation of canonical N-end rule substrates<sup>17</sup>.

To address this question, it was necessary to produce a variant of the Cup9p repressor that could be moderately overexpressed for pulse-chase analyses without influencing the expression of Ptr2p and the uptake of dipeptides. An Asn (N) → Ser (S) substitution at position 265 of Cup9p, within the recognition helix of the homeodomain<sup>19</sup>, was predicted to strongly reduce the affinity of Cup9p for DNA<sup>20</sup>. The N265S substitution did not alter the *in vivo* degradation of Cup9p (data

not shown). This Cup9p derivative, tagged at the C-terminus with the FLAG epitope and denoted as Cup9p<sub>NSF</sub>, was expressed as part of a fusion of the form <sup>F</sup>DHFR-Ub-Cup9p<sub>NSF</sub>, where <sup>F</sup>DHFR was the N-terminally FLAG-tagged mouse dihydrofolate reductase. Ub-specific proteases (UBPs) co-translationally cleave this UPR (Ub/protein/reference) fusion at the Ub-Cup9p junction, yielding the test protein Cup9p<sub>NSF</sub> and the long-lived <sup>F</sup>DHFR-Ub reference protein, which serves as an internal control for variations in expression levels and immunoprecipitation efficiency<sup>21,22</sup>.

Cells expressing Cup9p<sub>NSF</sub> were grown in minimal medium containing allantoin as the nitrogen source to avoid the known effects of nitrogen catabolite repression on *PTR2* expression<sup>23</sup>. Leu-Ala and Arg-Ala, dipeptides bearing either type of destabilizing N-terminal residue (Leu, bulky hydrophobic; Arg, basic), were added to the medium prior to pulse-chase analysis, to a final concentration of 10 mM (see Methods). This dipeptide concentration results in maximal inhibition of degradation of N-end rule substrates<sup>17</sup>. Strikingly, the addition of either Leu-Ala or Arg-Ala exerted an *opposite* effect on Cup9p<sub>NSF</sub>, strongly accelerating its degradation in wildtype (*UBR1*) cells. The half-life of Cup9p<sub>NSF</sub> decreased from ~5 min in the absence of dipeptides (Fig. 1c) to less than 1 min in their presence (Fig. 1b). This stimulatory effect was not observed in a *ubr1Δ* strain, indicating that the augmented degradation of Cup9p<sub>NSF</sub> was dependent on Ubr1p. The enhancement of degradation required dipeptides bearing destabilizing N-terminal residues: Ala-Leu and Ala-Arg, which bore a stabilizing N-terminal residue but had the same amino acid composition as, respectively, Leu-Ala and Arg-Ala, did not affect the degradation of Cup9p<sub>NSF</sub> ( $t_{1/2}$  ~5 min) (Fig. 1b and data not shown). Similar results were obtained with cells expressing Cup9<sub>NSF</sub> that was not a part of a UPR fusion (data not shown).

To determine the concentration dependence of the stimulation, we measured the degradation of Cup9p at a range of concentrations of Trp-Ala, another peptide bearing a hydrophobic (type 2) destabilizing N-terminal residue. The enhancement of Cup9p<sub>NSF</sub> degradation was detectable at 1 μM Trp-Ala, the lowest concentration tested ( $t_{1/2}$  ~1 min) (Fig. 1c). In contrast, the degradation of Cup9p<sub>NSF</sub> was not

altered either by Ala-Trp or by Trp and Ala, the free amino acid components of Trp-Ala and Ala-Trp (Fig. 1c). These results indicate that the relevant signaling molecule in this process is a dipeptide bearing a destabilizing N-terminal residue. Experiments with other dipeptides carrying destabilizing N-terminal residues (Leu-Ala and Arg-Ala) yielded similar results (data not shown). Thus, the range of dipeptide concentrations that significantly stimulated Cup9p degradation was similar to physiologically active levels of many other nutrients. Interestingly, these concentrations (1-10  $\mu$ M) were 1,000 to 10, 000-fold lower than the concentration (~10 mM) necessary for maximal inhibition of the Ubr1p-dependent degradation of protein substrates bearing N-degrons<sup>17</sup>. One possibility is that inhibiting N-degron-mediated proteolysis requires a dipeptide and a protein substrate to compete for the same binding site of Ubr1p, while no such competition exists when a dipeptide stimulates Cup9p degradation, since the dipeptide and Cup9p bind to two different sites of Ubr1p. It is also possible that only a small fraction of a cell's Ubr1p molecules have to be activated to enhance the degradation of Cup9p, while the Ubr1p pool may have to be almost saturated by dipeptides to suppress the targeting of a protein bearing an N-degron. Irrespective of the underlying mechanism, the above results, together with the observation that a protein bearing a destabilizing N-terminal residue can be co-expressed with Cup9p without altering the kinetics of Cup9p degradation (data not shown), suggest that cells are capable of independently regulating the Ubr1p-dependent degradation of proteins bearing N-degrons and internal degrons.

Cup9p represses transcription of the transporter-encoding *PTR2* gene<sup>4</sup>. Thus, the dipeptide-induced acceleration of Cup9p degradation would be expected to increase the levels of *PTR2* mRNA, ultimately leading to an increase in dipeptide uptake. In this circuit, dipeptides bearing destabilizing N-terminal residues act as effectors of a Ubr1p-based positive feedback loop, enhancing their own uptake by accelerating the Ubr1p-mediated degradation of Cup9p. This conjecture was tested by examining the levels of *PTR2* mRNA in the presence or absence of dipeptides in the medium. At 25  $\mu$ M, both Trp-Ala and Arg-Ala induced *PTR2* expression in the wildtype (*UBR1*) strain (Fig. 2a). The effect of Trp-Ala was particularly strong, resulting in levels of *PTR2* mRNA comparable to that of actin (*ACT1*) mRNA

(Fig. 2a). Both Ubr1p and Cup9p were required for these effects, since the expression of *PTR2* was not altered by dipeptides in *ubr1Δ* and *cup9Δ* strains. Testing a range of concentrations of Trp-Ala showed that induction of *PTR2* mRNA could be observed at 1 μM Trp-Ala, increased substantially at 10 μM Trp-Ala, and increased more gradually at higher concentrations, in agreement with the observed changes in the half-life of Cup9p at different levels of Trp-Ala (Fig. 1c).

A plausible mechanism of the enhancement effect is that a dipeptide interacts with either the basic or hydrophobic N-terminal residue-binding sites of Ubr1p, while a distinct (third) substrate-binding site of Ubr1p recognizes the internal degron of Cup9p. In this model, the interaction of Ubr1p with dipeptides allosterically increases the ability of the Ubr1p-Rad6p (E3-E2) complex to ubiquitylate Cup9p. To test whether dipeptides act directly through Ubr1p, we examined the effect of dipeptides on Cup9p ubiquitylation in an *in vitro* system consisting of the following purified components: Ubr1p (E3), Rad6p (E2), Uba1p (E1), Ub, ATP, and radiolabeled Cup9p. In this system, Cup9p was significantly multi-ubiquitylated, in a Ubr1p/Rad6p-dependent reaction (data not shown), in the absence of added dipeptides (Fig. 3). This result was consistent with the relatively rapid *in vivo* degradation of Cup9p ( $t_{1/2} \sim 5$  min) in the absence of dipeptides (Fig. 1c).

The addition of dipeptides bearing either type of destabilizing N-terminal residue to the *in vitro* system substantially stimulated the Ubr1p-dependent multi-ubiquitylation of Cup9p (Fig. 3). Dipeptides of the same composition but bearing a stabilizing N-terminal residue did not stimulate multi-ubiquitylation, nor did the amino acid components of these dipeptides (Fig. 3). Given the composition of this *in vitro* system, our results demonstrate that dipeptides act directly through Ubr1p, without an intermediate signaling pathway. The underlying allosteric mechanism may involve increased affinity of Ubr1p for Cup9p, or enhanced ubiquitylation activity of the Ubr1p-Rad6p complex towards Cup9, or both.

The findings of this work show that the two binding sites of Ubr1p that interact with destabilizing N-terminal residues have a specific physiological function as allosteric effector binding sites. These sites enable Ubr1p to sense the presence of imported dipeptides, and to induce the expression of the *Ptr2p* transporter accordingly, by accelerating degradation of the Cup9p repressor. The resulting

understanding of the regulation of peptide import is summarized in Figure 4a-d. This model predicts that a dipeptide bearing a destabilizing N-terminal residue, e. g., Leu-Ala, should stimulate its own uptake, in contrast to Ala-Leu, a dipeptide of the same composition bearing a stabilizing N-terminal residue. This prediction was borne out when we tested the ability of these two leucine-containing dipeptides to support the growth of *S. cerevisiae* auxotrophic for leucine. This Leu<sup>-</sup> strain grew robustly on plates supplemented with 230 μM Leu-Ala, but it did not form visible colonies in the presence of 230 μM Ala-Leu (Fig. 4e).

Food sources that *S. cerevisiae* encounters outside the laboratory setting are likely to contain mixtures of short peptides, a subset of which would be capable of activating the Ubr1p-based positive feedback circuit, since 12 of the 20 amino acid residues are destabilizing in the yeast N-end rule<sup>12</sup>. We modeled this situation by providing cells with a mixture of Leu-Ala at a low concentration (2 μM) and Ala-Leu at a high concentration (230 μM). Although neither dipeptide supplement alone could satisfy the strain's requirement for leucine, a mixture of the two dipeptides supported robust growth (Fig. 4e). Moreover, a mixture of Ala-Leu (230 μM) and a dipeptide *lacking* leucine but bearing a destabilizing N-terminal residue (10 μM Arg-Ala or 1 μM Lys-Ala) also rescued growth. In contrast, leucine-lacking dipeptides bearing stabilizing N-terminal residues, or the amino acid components of these dipeptides, could not rescue the growth of Leu<sup>-</sup> cells in the presence of 230 μM Ala-Leu (Fig. 4e). These results are predicted by the model described in Fig. 4a-d, in that a dipeptide bearing a destabilizing residue is expected to accelerate the import of all dipeptides, thereby enabling the uptake of sufficient Ala-Leu to satisfy the strain's leucine requirement.

This work establishes for the first time that the activity of an E3 can be directly linked to the presence of an environmental signal through an allosteric interaction with a small compound. Specifically, dipeptides bearing destabilizing N-terminal residues are shown to act as allosteric activators of Ubr1p, enhancing its ability to support the ubiquitylation and degradation of Cup9p. Physiologically, this results in a positive feedback circuit governing the uptake of peptides. By binding to Ubr1p, the imported dipeptides accelerate degradation of Cup9p, thereby derepressing the

synthesis of the Ptr2p transporter and enhancing the cell's ability to import di- and tripeptides.

Most eukaryotic cells have the capacity to import peptides. The sequence conservation of both the basic and hydrophobic N-terminal residue-binding sites between the mouse and yeast Ubr1p proteins<sup>24</sup> suggests that the peptide-sensing physiological functions of these sites may be conserved among eukaryotes. In metazoans, many neurotransmitters and hormones bear destabilizing N-terminal residues, so it is possible that some of these signaling molecules act at least in part intracellularly, by modulating Ubr1p-mediated protein degradation. The ClpAP-dependent N-end rule pathway of *E. coli*<sup>25</sup>, whose physiological functions are unknown, may also play a role in regulating the import of peptides or related compounds, by analogy with the *S. cerevisiae* N-end rule pathway.

The discovery of peptide-mediated positive feedback (Fig. 4) identifies the physiological rationale for control of peptide import through the Ubr1p-dependent degradation of Cup9p: the N-terminal residue-binding sites of Ubr1p enable it to sense intracellular dipeptide concentrations and modulate the degradation of Cup9p accordingly. In addition, our results expand the set of mechanisms that regulate protein degradation from the previously known modulatory phosphorylation of degrons and E3's to allosteric regulation of an E3 by a small compound. The Ub system is either known or suspected to play major roles in the control of intermediary metabolism and the transport of small molecules across membranes<sup>2,3</sup>. Our findings suggest that these compounds, or their enzymatically produced derivatives, may modulate the functions of E3s in the Ub system similarly to the effects observed here with dipeptides and Ubr1p.

## Methods

**Yeast strains and plasmids.** The *S. cerevisiae* strains used in pulse-chase experiments, JD52 (*MAT $\alpha$  lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*), JD55 (*ubr1Δ::HIS3*), were described previously<sup>26</sup>.  $\text{p}_{\text{DHFR}}\text{-Ub-Cup9p}_{\text{NSF}}$  was expressed from the  $\text{P}_{\text{MET25}}$  promoter on the centromeric vector p416MET25<sup>27</sup>. Construction details are available upon request. Strains used for Northern analyses were AVY30 (*MAT $\alpha$  leu2-3,112 ubr1Δ::LEU2*), AVY 31 (*MAT $\alpha$  leu2-3,112 cup9Δ::LEU2*), and AVY32 (*MAT $\alpha$  LEU2*), constructed in the RJD350 background (*MAT $\alpha$  leu2-3,112*; a gift from R. Deshaies) using restriction fragments obtained from plasmids pSOB30<sup>5</sup>, pCB119<sup>4</sup> and pJJ252<sup>28</sup>, respectively. The RJD350 strain was used for the colony formation assay (Fig. 4e). SHM plates<sup>4</sup> were supplemented with dipeptides or amino acids at the following concentrations: 230 μM Leu-Ala (or 2 μM where noted), 230 μM Ala-Leu, 230 μM each of Leu and Ala, 10 μM Arg-Ala, 10 μM Ala-Arg, 1 μM Lys-Ala, 1 μM Ala-Lys, 10 μM each of Arg and Ala, or 1 μM each of Lys and Ala.

**Pulse-chase analysis.** Cells were cultured in SHM<sup>4</sup> with auxotrophic supplements. Dipeptides were added to cultures at  $\text{OD}_{600} \sim 0.6$  and incubation continued for 2.5 h in the experiments in Fig 1b, and for 30 min for Fig 1c. Cells were harvested, washed in 0.8 ml of SHM, resuspended in 0.4 ml of SHM, and labeled for 5 min at 30°C with 0.16 mCi of  $^{35}\text{S}$ -EXPRESS (New England Nuclear). Cells were pelleted, and resuspended in fresh SHM containing 4 mM L-methionine and 2 mM L-cysteine. 0.1 ml samples were taken at the time points indicated and transferred to chilled tubes, each containing 0.5 ml of 0.5-mm glass beads, 0.7 ml of ice-cold lysis buffer (1% Triton-X100, 0.15 M NaCl, 5 mM EDTA, 50 mM Na-HEPES, pH 7.5), and a mixture of protease inhibitors (final concentrations 1 mM phenylmethylsulfonyl fluoride, 2 μg/ml aprotinin, 0.5 μg/ml leupeptin, and 0.7 μg/ml pepstatin). Extracts were prepared and immunoprecipitations carried out as described<sup>26</sup>, using anti-FLAG M2 resin (Sigma). Immunoprecipitates were fractionated by SDS-13% PAGE, and detected by autoradiography.

**RNA preparation and Northern analysis.** Cells were cultured in SHM to  $\text{OD}_{600} \sim 0.6$ . The indicated dipeptides were then added to a 25 μM final concentration, and the incubation was continued for an additional 30 min. Total

RNA was prepared<sup>29</sup>, and 25 µg samples were electrophoresed in 1% formaldehyde-agarose gels, followed by blotting for Northern analysis<sup>30</sup>.

**Ubr1p-dependent *in vitro* ubiquitylation system.** The components of this system were purified as follows. N-terminally hexahistidine-tagged Uba1p was overexpressed in *S. cerevisiae* and purified by fractionation over Ni-NTA, ubiquitin affinity, and Superdex-200 columns. Rad6p was overexpressed in *E. coli*, and purified by fractionation over DEAE, Mono-Q and Superdex-75 columns. N-terminally FLAG-tagged Ubr1p was overexpressed in *S. cerevisiae*, and purified by fractionation over anti-FLAG M2, Ubc2p affinity, and Superdex-200 columns. N-terminally FLAG-tagged, C-terminally hexahistidine-tagged Cup9p was expressed and radiolabeled in *E. coli*, and purified by fractionation over Ni-NTA and anti-FLAG M2 columns. Details of these expression and purification protocols are available upon request.

The *in vitro* ubiquitylation reactions contained the following components: 7 µM Ub, 50 nM Uba1p, 50 nM Rad6p, 50 nM Ubr1p, 550 nM <sup>35</sup>S-labeled Cup9p, 25 mM HEPES/KOH (pH 7.5), 25 mM KCl, 5 mM MgCl<sub>2</sub>, 2 mM ATP, 0.1 mM dithiothreitol, and 0.5 mg/ml ovalbumin as a carrier protein. Dipeptides or amino acids, as indicated, were added to a final concentration of 2 µM (top panel, Fig. 3) or 10 µM (bottom panel, Fig. 3). All components except Uba1p were mixed on ice for 10 minutes; Uba1p was then added and reactions shifted to 30°C. After the indicated times, the reactions were terminated by adding an equal volume of 2 x SDS-PAGE loading buffer and heating at 95°C for 5 min, followed by 8% SDS-PAGE.

### Acknowledgements

We thank A. Webster for her valuable contributions to the establishment of the *in vitro* system, and R. Deshaies, J. Dohmen, L. Prakash, H. Rao and J. Sheng for their gifts of plasmids and strains. We also thank C. Byrd, H. Rao, and especially R. Deshaies for helpful discussions and comments on the manuscript. This work was supported by grants to A. V. from the NIH. Correspondence and requests for materials should be addressed to A. V. (e-mail: avarsh@caltech.edu).

**References:**

1. Laney, J. D. & Hochstrasser, M. Substrate targeting in the ubiquitin system. *Cell* **97**, 427-430 (1999).
2. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **76**, 425-479 (1998).
3. Varshavsky, A. The ubiquitin system. *Trends Biochem. Sci.* **22**, 383-387 (1997).
4. Byrd, C., Turner, G. C. & Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* **17**, 269-277 (1998).
5. Bartel, B., Wünnigen, I. & Varshavsky, A. The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179-3189 (1990).
6. Scheffner, M., Smith, S. & Jentsch, S. in *Ubiquitin and the Biology of the Cell* (eds. Peters, J.-M., Harris, J. R. & Finley, D.) 65-98 (Plenum Press, New York, 1998).
7. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367-380 (1998).
8. King, R. W., Deshaies, R. J., Peters, J. M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274**, 1652-1659 (1996).
9. Kotani, S., Tanaka, H., Yasuda, H. & Todokoro, K. Regulation of APC activity by phosphorylation and regulatory factors. *J. Cell Biol.* **146**, 791-800 (1999).
10. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186 (1986).
11. Bachmair, A. & Varshavsky, A. The degradation signal in a short-lived protein. *Cell* **56**, 1019-1032 (1989).
12. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**, 12142-12149 (1996).
13. Xie, Y. M. & Varshavsky, A. The N-end rule pathway is required for import of histidine in yeast lacking the kinesin-like protein Cin8p. *Current Genetics* **36**, 113-123 (1999).
14. Dohmen, R. J., Madura, K., Bartel, B. & Varshavsky, A. The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. *Proc. Natl. Acad. Sci. USA* **88**, 7351-7355 (1991).

15. Reiss, Y., Kaim, D. & Hershko, A. Specificity of binding of N-terminal residues of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.* **263**, 2693-2269 (1988).
16. Gonda, D. K. *et al.* Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700-16712 (1989).
17. Baker, R. T. & Varshavsky, A. Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **87**, 2374-2378 (1991).
18. Alagramam, K., Naider, F. & Becker, J. M. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **15**, 225-234 (1995).
19. Knight, S. A. B., Tamai, K. T., Kosman, D. J. & Thiele, D. J. Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.* **14**, 7792-7804 (1994).
20. Wolberger, C., Vershon, A. K., Liu, B. S., Johnson, A. D. & Pabo, C. O. Crystal-Structure of a Mat Alpha-2 Homeodomain-Operator Complex Suggests a General-Model For Homeodomain-DNA Interactions. *Cell* **67**, 517-528 (1991).
21. Lévy, F., Johnsson, N., Rumenapf, T. & Varshavsky, A. Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907-4912 (1996).
22. Suzuki, T. & Varshavsky, A. Degradation signals in the lysine-asparagine sequence space. *EMBO J.* **18**, 6017-6026 (1999).
23. Barnes, D., Lai, W., Breslav, M., Naider, F. & Becker, J. M. PTR3, a novel gene mediating amino acid-inducible regulation of peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **29**, 297-310 (1998).
24. Kwon, Y. T. *et al.* The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **95**, 7898-7903 (1998).
25. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science* **254**, 1374-1377 (1991).
26. Ghislain, M., Dohmen, R. J., Levy, F. & Varshavsky, A. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 4884-4899 (1996).

27. Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae* - comparison of transcriptional activity and their use for heterologous expression. *Nucl. Acids Res.* **22**, 5767-5768 (1994).
28. Jones, J. S. & Prakash, L. Yeast *Saccharomyces cerevisiae* selectable markers in pUC18 polylinkers. *Yeast* **6**, 363-366 (1990).
29. Schmitt, M. E., Brown, T. A. & Trumppower, B. L. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucl. Acids Res.* **18**, 3091-3092 (1990).
30. Ausubel, F. M. *et al.* (eds.) *Current Protocols in Molecular Biology*. (Wiley-Interscience, New York, 1996).

## Figure Legends

**Figure 1.** Enhancement of Cup9p degradation by dipeptides bearing destabilizing N-terminal residues. **a**, The fusion protein used for pulse-chase analysis. The stable  ${}^3\text{F}$ DHFR-Ub reference portion of the fusion is co-translationally cleaved from Cup9<sub>NSF</sub> by UBP<sub>s</sub>. **b**, Pulse-chase analysis of  ${}^3\text{F}$ DHFR-Ub-Cup9<sub>NSF</sub> in the presence of various dipeptides at 10 mM. Dipeptides bearing either basic (Arg-Ala) or bulky hydrophobic (Leu-Ala) destabilizing N-terminal residues strongly enhance Cup9<sub>NSF</sub> degradation, but only in strains expressing Ubr1p. Dipeptides bearing a stabilizing N-terminal residue (Ala-Arg and Ala-Leu) do not alter Cup9<sub>NSF</sub> degradation. **c**, The effects of different concentrations of Trp-Ala on the enhancement of Cup9<sub>NSF</sub> degradation in wildtype (*UBR1*) cells. Lanes marked by a dash indicate pulse-chase analysis performed in the absence of added dipeptides. Enhancement of Cup9<sub>NSF</sub> degradation was detectable at 1  $\mu\text{M}$  Trp-Ala, and became substantially greater at 10  $\mu\text{M}$ .

**Figure 2.** Effects of dipeptides on expression of the dipeptide transporter gene *PTR2*. **a**, Induction of *PTR2* expression by dipeptides bearing destabilizing N-terminal residues (Trp-Ala and Arg-Ala) required both *UBR1* and *CUP9*. Dipeptides bearing a stabilizing N-terminal residue (Ala-Trp and Ala-Arg) had no effect on *PTR2* expression. *PTR2* mRNA and the *ACT1* mRNA loading control are indicated. **b**, Effect of different concentrations of Trp-Ala on the levels of *PTR2* mRNA.

**Figure 3.** *In vitro* ubiquitylation of Cup9p is enhanced by dipeptides bearing destabilizing N-terminal residues. Reactions consisting of purified Uba1p (E1), Rad6p (E2), Ubr1p (E3), Ub, ATP and radiolabeled Cup9p, were supplemented with the indicated dipeptides or amino acids (top panel, 2  $\mu\text{M}$ , bottom panel, 10  $\mu\text{M}$ ), or left un-supplemented (denoted by a dash), and allowed to proceed at 30°C for the designated times. Radiolabeled input Cup9p, and its multi-ubiquitylated derivatives are indicated.

**Figure 4.** Feedback regulation of peptide import in *S. cerevisiae*. **a**, Genetic diagram of the peptide transport circuit. **b**, Ubr1p is required for dipeptide uptake. In the absence of Ubr1p (*ubr1Δ*), the transcriptional repressor Cup9p is long-lived, accumulates to high levels, and extinguishes the expression of the *PTR2* gene. Thus, *ubr1Δ* cells cannot import dipeptides (red dots). **c**, In a wildtype (*UBR1*) cell growing in the absence of extracellular dipeptides, Ubr1p targets Cup9p for degradation ( $t_{1/2} \sim 5$  min), resulting in a moderate concentration of Cup9p and weak but significant expression of the *Ptr2p* transporter (blue double ovals). **d**, In wildtype (*UBR1*) cells growing in the presence of extracellular dipeptides some of which bear destabilizing N-terminal residues, the imported dipeptides bind to either the basic (type 1) or the hydrophobic (type 2) residue-binding site of Ubr1p. These peptides are denoted as a red block and a green wedge, respectively. Binding of either type of dipeptide to Ubr1p allosterically increases the rate of Ubr1p-mediated degradation of Cup9p. Peptides of both types are shown as bound to Ubr1p, but in fact the binding of either peptide accelerates Cup9p degradation. The resulting decrease of the half-life of Cup9p from ~5 min to less than 1 min results in a very low concentration of Cup9p, and consequently to a strong induction of the *Ptr2p* transporter. **e**, Colony formation assays. A *S. cerevisiae* strain requiring leucine for growth was incubated on plates supplemented with either dipeptides or their amino acid constituents at the following concentrations: Leu-Ala, 230  $\mu$ M (or 2  $\mu$ M, as indicated); Ala-Leu, 230  $\mu$ M; Arg-Ala, 10  $\mu$ M; Ala-Arg, 10  $\mu$ M; Lys-Ala, 1  $\mu$ M; Ala-Lys, 1  $\mu$ M; Arg and Ala, 10  $\mu$ M each, Lys and Ala, 1  $\mu$ M each.

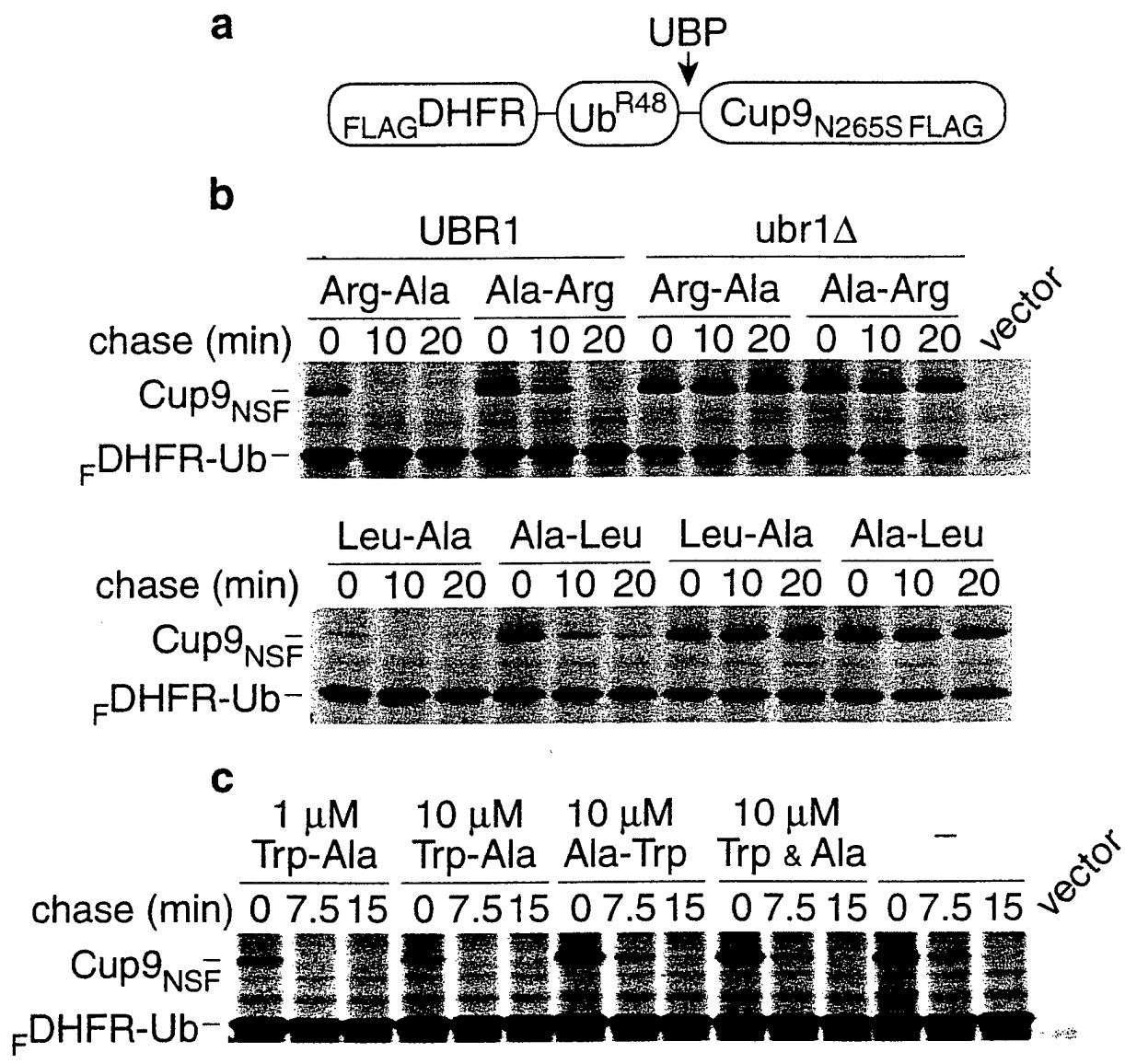


FIGURE 1

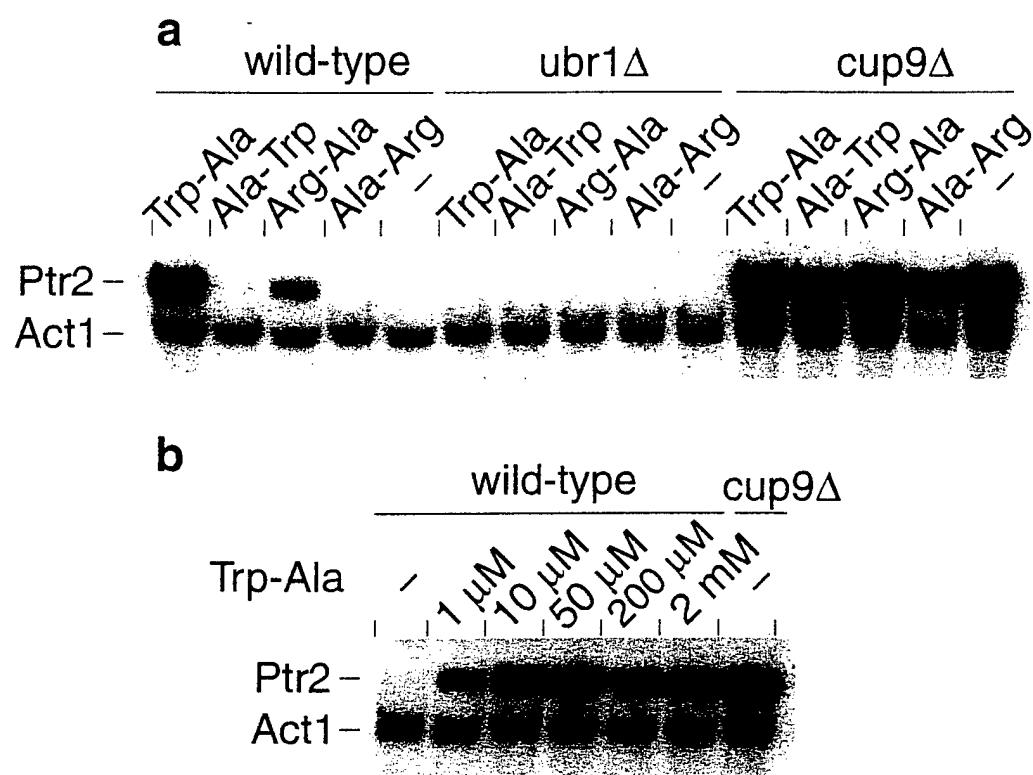
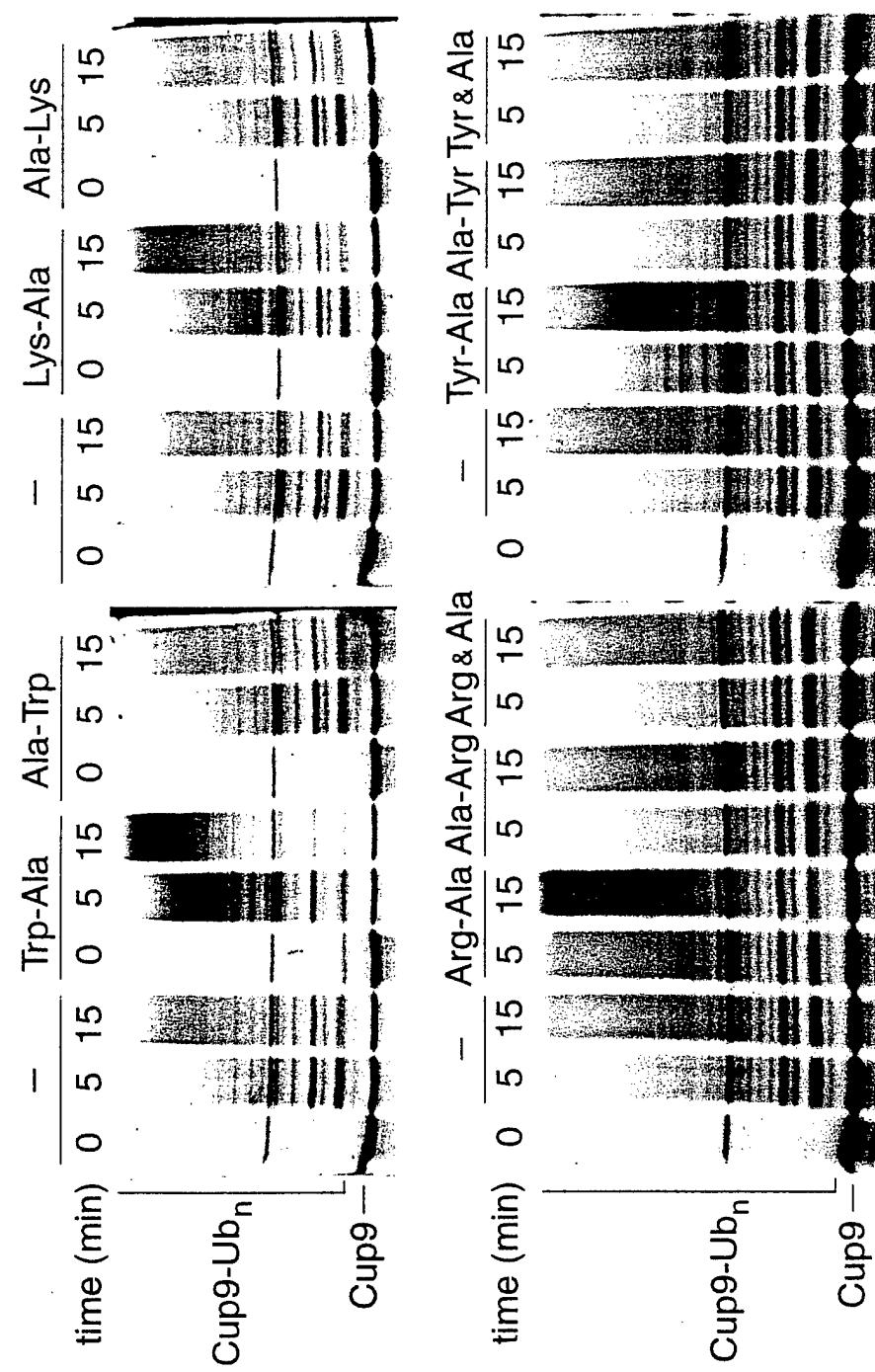


FIGURE 2

FIGURE 3



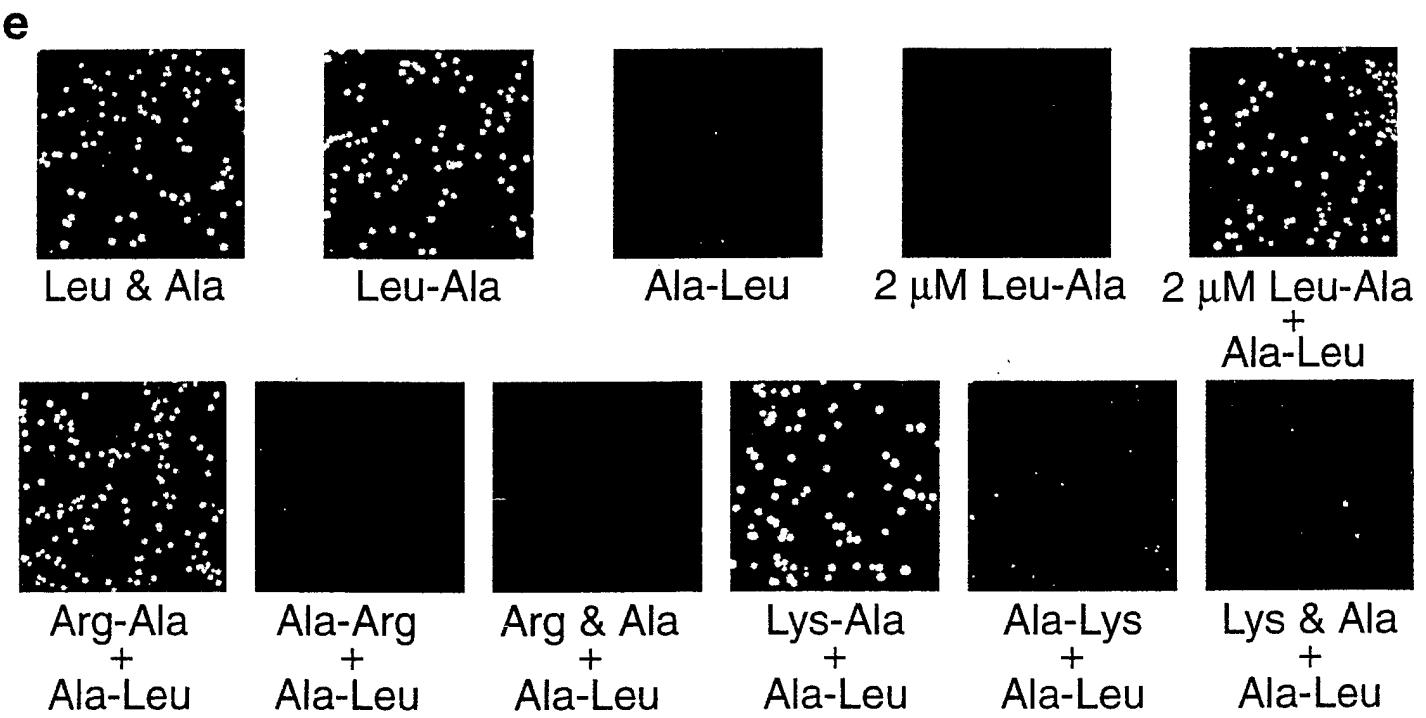
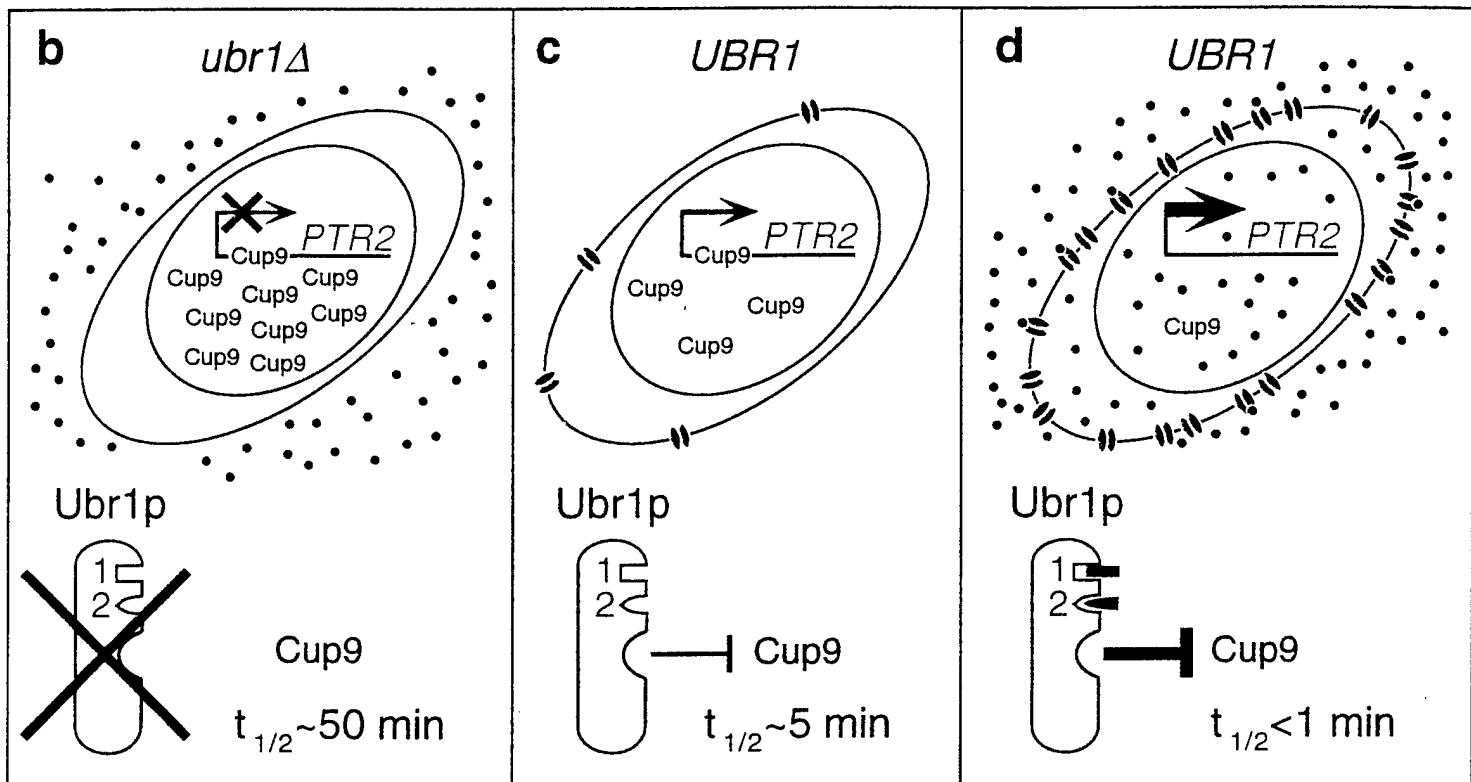
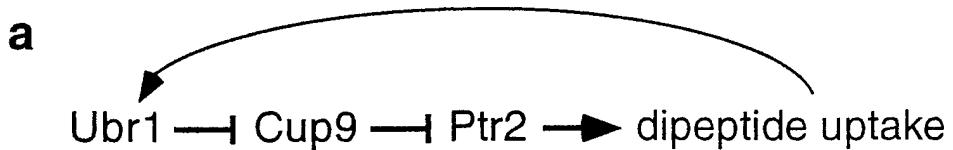


FIGURE 4

# **Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway**

Yong Tae Kwon<sup>1</sup>, Seth A. Balogh<sup>2</sup>, Ilia V. Davydov<sup>1</sup>, Anna S. Kashina<sup>1</sup>,  
Jeong Kyo Yoon<sup>1</sup>, Youming Xie<sup>1</sup>, Arti Gaur<sup>1,3</sup>, Lynn Hyde<sup>2,4</sup>, Victor H. Denenberg<sup>2,5</sup>,  
and Alexander Varshavsky<sup>1\*</sup>

<sup>1</sup>Division of Biology, California Institute of Technology, Pasadena, CA 91125.

<sup>2</sup>Biobehavioral Sciences Graduate Program, University of Connecticut,  
Storrs, CT 06269-4154.

<sup>3</sup>Present address: Institute for Genetics, University of Cologne, Cologne,  
D-50931, Germany.

<sup>4</sup>Present address: Department of Pediatrics and Psychiatry, University of Colorado  
School of Medicine, Denver, CO 80262.

<sup>5</sup>Department of Psychology, University of Connecticut, Storrs, CT 06269-4154.

\*Corresponding author: Alexander Varshavsky, Division of Biology, 147-75, Caltech,  
1200 East California Blvd., Pasadena, CA 91125.

Telephone: 626-395-3785; Fax: 626-440-9821; E-mail: avarsh@caltech.edu

*Running title:* Mouse N-terminal Amidase

*Key words:* / N-terminal amidase / ubiquitin / proteolysis / mouse / behavior /  
/ N-end rule / gene targeting / spatial learning /

**ABSTRACT**

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. N-terminal asparagine and glutamine are tertiary destabilizing residues, in that they are enzymatically deamidated to yield secondary destabilizing residues aspartate and glutamate, which are conjugated to arginine, a primary destabilizing residue. N-terminal arginine of a substrate protein is bound by the *Ubr1*-encoded E3 $\alpha$ , the E3 component of the ubiquitin/proteasome-dependent N-end rule pathway. We describe the construction and analysis of mouse strains lacking the asparagine-specific N-terminal amidase (Nt<sup>N</sup>-amidase), encoded by the *Ntan1* gene. In wildtype embryos, *Ntan1* was strongly expressed in the branchial arches, and in the tail and limb buds. The *Ntan1*<sup>-/-</sup> mouse strains lacked the Nt<sup>N</sup>-amidase activity but retained glutamine-specific Nt<sup>Q</sup>-amidase, indicating that the two enzymes are encoded by different genes. Among the normally short-lived N-end rule substrates, only those bearing N-terminal asparagine became long-lived in *Ntan1*<sup>-/-</sup> fibroblasts. The *Ntan1*<sup>-/-</sup> mice were fertile and outwardly normal, but differed from their congenic wildtype counterparts in spontaneous activity, spatial memory, and a socially conditioned exploratory phenotype that has not been previously described with other mouse strains.

## INTRODUCTION

A multitude of regulatory circuits involve conditionally or constitutively short-lived proteins (26, 27, 44, 48, 49, 64). Features of proteins that confer metabolic instability are called degradation signals, or degrons (37, 63). The essential component of one degradation signal, termed the N-degron, is a destabilizing N-terminal residue of a protein (3). A set of N-degrons containing different N-terminal residues which are destabilizing in a given cell yields a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. An N-end rule pathway is present in all organisms examined, from mammals and plants to fungi and prokaryotes (63).

In eukaryotes, an N-degron comprises two determinants: a destabilizing N-terminal residue and an internal lysine of a substrate protein (4, 32, 60). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (15, 49). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system (25-27). Ub is a 76-residue eukaryotic protein that exists in cells either free or covalently conjugated to many other proteins. The Ub system plays a role in a vast range of processes, including cell growth, division, differentiation, and responses to stress. In most of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (10, 17, 20, 51).

The N-end rule has a hierachic structure. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded<sup>†</sup> N-terminal amidohydrolase (Nt-amidase), into the secondary destabilizing N-terminal residues Asp and Glu (6). Destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *S. cerevisiae* *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase) (8, 41), to Arg, one of the primary destabilizing residues (Fig. 1A). In mammals, the deamidation step is mediated by two Nt-amidases, Nt<sup>N</sup>-amidase and Nt<sup>Q</sup>-amidase, which are specific, respectively, for N-terminal Asn and Gln (Fig. 1A) (24, 59). The mammalian counterpart of the yeast R-transferase *Ate1p* exists as two distinct species, *ATE1-1p* and *ATE1p-2*, which are produced through alternative splicing of *Ate1* pre-mRNA (34).

In vertebrates, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (Fig. 1A) (18, 23). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognition (E3) component of the N-end rule pathway. In *S. cerevisiae*, Ubr1p is a 225 kDa protein which recognizes potential N-end rule substrates through its type 1 and type 2 substrate-binding sites. The type 1 site binds the basic N-terminal residues Arg, Lys, and His. The type 2 site binds the bulky hydrophobic N-terminal terminal residues Phe, Leu, Trp, Tyr, and Ile (35, 63). Ubr1p contains yet another substrate-binding site that targets proteins such as Cup9p and Gpa1p, which bear internal (non-N-terminal) degrons (12, 54). The *Ubr1* genes encoding mouse and human N-recognition, also called E3 $\alpha$ , have been cloned (36), and mouse strains lacking *Ubr1* have recently been constructed (Y. T. Kwon and A. Varshavsky, unpublished data).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2* which encodes the peptide transporter (1, 12); a mechanistically undefined role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (47); the degradation of alphaviral RNA polymerases and other viral proteins in infected metazoan cells (19, 38); and the degradation of Gpa1p, a G $\alpha$  protein of *S. cerevisiae* (43, 54). Physiological N-end rule substrates were also identified among the proteins secreted into the mammalian cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes*. Short half-lives of these bacterial proteins are required for the efficient presentation of their peptides to the immune system (56). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (28), and to delay limb regeneration in amphibians (61). Studies of the Ub-dependent proteolysis of endogenous proteins in muscle extracts suggested that the N-end rule pathway plays a role in catabolic states that result in muscle atrophy (39, 57). A crush injury to the rat sciatic nerve was reported to result in a ~10-fold increase in the rate of arginine conjugation to the N-termini of proteins in the nerve's region upstream of the crush site, suggesting an

injury-induced increase in the concentration of R-transferase substrates and/or an enhanced activity of the N-end rule pathway (65).

Physiological substrates of either yeast or metazoan Nt-amidases and R-transferases are unknown. Engineered N-end rule substrates, including substrates of Nt-amidases and R-transferases, can be produced *in vivo* through the Ub fusion technique, in which a Ub-X-reporter fusion is cleaved by deubiquitylating enzymes (DUBs) (66) after the last residue of Ub, yielding a reporter bearing the desired N-terminal residue X (3, 63).

The mouse Asn-specific Nt<sup>N</sup>-amidase is encoded by the 17 kb *Ntan1* gene. The 1.6 kb *Ntan1* mRNA specifies the 310-residue Nt<sup>N</sup>-amidase (24). In the present work, we characterized the expression and intracellular localization of Nt<sup>N</sup>-amidase. We also constructed mouse strains bearing a homozygous deletion/disruption of *Ntan1*, and showed that these mice lacked both Nt<sup>N</sup>-amidase and the Asn-specific branch of the N-end rule pathway. The *Ntan1*<sup>-/-</sup> mice were fertile, outwardly normal, but were found to differ from their congenic wildtype counterparts in their spontaneous activity and spatial memory. Among these differences was a socially conditioned exploratory phenotype of *Ntan1*<sup>-/-</sup> mice that has not been previously described, to our knowledge, with other mouse strains.

## MATERIALS AND METHODS

**Construction of mouse strains lacking Nt<sup>N</sup>-amidase.** Genomic *Ntan1* DNA fragments were subcloned from the P1 phage DNA containing the strain C129-derived mouse *Ntan1* gene (24). The 1,462 bp *Hind*III-*Pst*I fragment containing exon 6 and its flanking introns was used as the vector's short homology arm. The 7.2 kb *Xho*I-*Not*I fragment containing exon 1 and its flanking intron was used as the long homology arm. To construct the targeting vector, a 1,462 bp *Hind*III-*Pst*I fragment containing exon 6 and its flanking introns (the vector's short arm) and the PGK/TK cassette of pPNT (a gift from R. C. Mulligan, Harvard Medical School, Boston, MA) were inserted into pPGKRN containing the wildtype PGK/neo cassette (a gift from R. Jaenisch, Whitehead Institute, Cambridge, MA), yielding pPGK-SA. The vector's 7.2 kb long arm was produced by joining, within pPGK-SA, a 4.8 kb *Xho*I-*Eco*RI

fragment to its flanking 2.6 kb *Eco*RI-*Eco*RI fragment containing exon 1 and flanking intron pPGK-SA, yielding the *Ntan1* targeting vector pNTAN1-KO (Fig. 1B). The *Xba*I-linearized targeting vector (Fig. 1B) was electroporated into CJ7 ES cells. Selection with G418 (at 0.4 mg/ml) and 1-(2'-deoxy, 2' fluoro-β-d-arabinofuranosyl)-5-iodouracil (FIAU) (at 0.4 μM) was started 24 h after electroporation. Correctly targeted ES cells were identified by PCR and Southern hybridization with the 5' and 3' probes (Fig. 1C, D). Cells of the 12 independent ES cell clones were injected into 3.5 dpc (days postcoitum) C57BL/6J blastocysts. The resulting male chimeras were bred with C57BL/6J females to test for germline transmission of the mutated *Ntan1* gene. The *Ntan1*<sup>+/−</sup> mice resulting from this cross (6 out of 12 independent ES clones were found to populate germline in these tests) were intercrossed to produce *Ntan1*<sup>−/−</sup> mice. Alternatively, the initial male chimeras were mated with 129/SvEv females, yielding, through the analogous series of steps, *Ntan1*<sup>−/−</sup> mice in the strain-129 background. All of the behavioral tests were carried out with *Ntan1*<sup>−/−</sup> mice in the strain-129 background. For genotyping, the tail-derived DNA was analyzed by PCR, or digested with either *Bam*HI or *Hind*III/*Xba*I, and analyzed by Southern hybridization. A 0.75 kb *Pst*I fragment and a 0.9 kb *Pst*I fragment (indicated in Fig. 1B) were used as the 5' and 3' hybridization probes, respectively.

**Northern and RT-PCR analyses of RNA.** Total RNA was isolated from brain, testis and EF cells of wildtype and *Ntan1*<sup>−/−</sup> mice as described (34). RNA was fractionated by electrophoresis in 1% formaldehyde-agarose gels, blotted onto Hybond N<sup>+</sup> (Amersham), and hybridized with <sup>32</sup>P-labeled probes specific for different regions of the *Ntan1* cDNA (GenBank #U57692). Probe a, nt 34-900; probe b, nt 118-450; probe c, nt 118-900; probe d, nt 34-450; probe e, nt 470-670; probe f, nt 680-900. Alternatively, RT-PCR was carried out (2), with first-strand cDNA synthesized using Superscript II polymerase (GIBCO, Frederick, MD). PCR was done using primers specific for different regions of the *Ntan1* cDNA. Primer P1, 5'-ATGCCACTGCTGGTGGATGG-GCAG (forward); P2, 5'-GAGCCAGACTTCTCAGAGGTCAG (forward); P3, 5'-GACA-TTCACTTAGTGACATTATG (forward); P4, 5'-TTCTGTGACAGCTGCCTGTCATC (reverse); P5, 5'-CATCAAGGTAGATCTAATATGTT (reverse). The ratio of mouse *Ate1*-1 and *Ate1*-2 mRNAs was determined as described (34).

**Whole-mount *in situ* hybridization.** Wildtype and *Ntan1*<sup>-/-</sup> embryos were staged, fixed, and processed for *in situ* hybridization as described (16). A 0.3 kb fragment of the *Ntan1* cDNA (nt 108-448) that was encompassed by the deleted region in the *Ntan1*<sup>-/-</sup> allele was subcloned into *Xba*I-*Xho*I sites of Bluescript II SK<sup>+</sup> (Stratagene), and the resulting plasmid, pMR27, was used as a template for synthesizing antisense RNA probe labeled with digoxigenin (Roche Molecular Biochemicals, Indianapolis, IN).

**Localization of NTAN1p-GFP.** The mouse *Ntan1* ORF was subcloned into the *Xho*I and *Age*I sites of pEGFP-N1 (Clontech), yielding pNTAN1-GFP, which expressed the NTAN1p-GFP fusion protein from the P<sub>CMV</sub> promoter. NIH 3T3 cells (ATCC 1658-CRL) were grown as monolayers in Dulbecco's Modified Eagle Medium (Gibco-BRL) supplemented with 10% fetal bovine serum. Cells were grown to ~15% confluence on glass coverslips for 24 hr prior to transfection with either the GFP-expressing control vector pEGFP-N1 or pNTAN1-GFP, using Lipofectamine (GIBCO) and the manufacturer-supplied protocol. Cells were incubated for 5 hr at 37°C in the serum-free media containing DNA and Lipofectamine. An equal volume of media containing 20% serum was then added, and the cells were grown for another 12 to 20 hr at 37°C. Cells were fixed with 2% formaldehyde in phosphate-buffered saline, and GFP fluorescence was examined using Zeiss Axiophot microscope.

**Immortalization of embryonic fibroblasts, transfection, and pulse-chase analysis.** Embryonic fibroblast (EF) cells were isolated from either *Ntan1*<sup>-/-</sup> or +/+ 13.5 days old (e13.5) embryos as described (52). EFs were grown in DMEM/F12 medium (GIBCO) supplemented with 15% fetal bovine serum, antibiotics, and 2 mM L-glutamine. They were immortalized using the SV40 virus (14). The SV40-transformed cell lines were transiently transfected, using Lipofectamine PLUS (GIBCO), with the plasmids pRC/dhaUbXnsP4 $\beta$ gal, which expressed DHFR-ha-Ub<sup>R48</sup>-X-nsP4 $\beta$ gal test proteins (X=Met, Asn, Gln, or Arg) (40) (see the main text). Cells were labeled with <sup>35</sup>S-Translabel (New England Nuclear, Boston, MA), followed by a chase for 0, 1, and 2 hr in the presence of cycloheximide, preparation of extracts, immunoprecipitation, SDS-PAGE, autoradiography, and quantitation with PhosphorImager, essentially as described (40).

**X-DHFR test proteins.**  $^{35}\text{S}$ -labeled X-DHFR proteins (X = Asn, Gln, or Asp) were prepared as described (24). Briefly, the plasmids pSG4, pSG41, and pSG44 expressing, respectively, Ub-Asn-DHFR, Ub-Asp-DHFR and Ub-Gln-DHFR from the  $P_{\text{trc}}$  promoter (24) were transformed into *E. coli* JM101 carrying pJT184, which expressed Ubp1p, a deubiquitylating enzyme (DUB) of *S. cerevisiae* (62). Cells of a 50-ml culture grown at 37°C to an  $A_{600}$  of ~0.9 in LB plus ampicillin (40  $\mu\text{g}/\text{ml}$ ) and chloramphenicol (20  $\mu\text{g}/\text{ml}$ ), were pelleted and resuspended in 50 ml of M9 supplemented with glucose (0.2%), thiamine (2  $\mu\text{g}/\text{ml}$ ), ampicillin (40  $\mu\text{g}/\text{ml}$ ), 1 mM isopropylthio- $\beta$ -D-galactoside (IPTG), and Methionine Assay Medium (Difco). The suspension was shaken for 1 hr at 37°C, followed by the addition of 1 mCi of  $^{35}\text{S}$ -Translabel (ICN) and further incubation for 1 hr at 37°C. Unlabeled L-methionine was then added to 1 mM, and the shaking was continued for another 10 min. The cells were harvested and the lysate were prepared by the addition of lysozyme and Triton X-100, followed by centrifugation (24). The supernatant, containing [ $^{35}\text{S}$ ]X-DHFR (X=Asn, Asp, or Glu) was purified by affinity chromatography on a methotrexate column (Pierce) (0.5 ml bed volume). [ $^{35}\text{S}$ ]X-DHFRs were examined by SDS-PAGE; they were greater than 95% pure.

**Isoelectrofocusing assay for amidase activity.** Extracts from wildtype or *Ntan1*<sup>-/-</sup> EF cells were prepared by homogenization in 0.01% Triton X-100, 10 % glycerol, 125 mM KCl, 7.5 mM MgCl<sub>2</sub>, 1 mM DTT, 0.25 mM EDTA, 0.2 mM phenylmethylsulfonyl fluoride (PMSF), 130 mM Tris-HCl (pH 7.5), and the protease inhibitors antipain, chymostatin, leupeptin, pepstatin and aprotinin (Sigma), each at 25  $\mu\text{g}/\text{ml}$ . For the deamidation assay, 5  $\mu\text{l}$  of  $^{35}\text{S}$ -labeled X-DHFR (0.5 mg/ml in storage buffer; X=Asn, Asp, or Glu) was mixed with 20  $\mu\text{l}$  of the extract, incubated for 2 hr at 37°C, and placed on ice. Samples (5  $\mu\text{l}$ ) were applied onto isoelectrofocusing polyacrylamide (IEF-PAGE) plates (pH 3.5-9.5) (Pharmacia, Piscataway, NJ, USA) precooled to 10°C. IEF was carried out for 80 min at 30 W in a cooled IEF apparatus (Hoefer, San Francisco, CA). The plates were soaked in 100 ml of 10% CCl<sub>3</sub>COOH, 5% 5-sulfosalicylic acid, stained with Coomassie Blue to detect IEF markers (Pharmacia), and autoradiographed.  $^{35}\text{S}$  in the bands of more acidic (deamidated) and more basic (initial) X-DHFR species was quantitated using PhosphorImager (Molecular

Dynamics, Sunnyvale, CA).

**Behavioral tests.** Mice were kept on a 0600- to 1800-hr light cycle and tested during the light period. For the rotarod, weight retention, coat-hanger and platform-leaving tests, the strain-129 *Ntan1*<sup>-/-</sup> mice and their congenic +/+ littermates (produced through matings of *Ntan1*<sup>+/+</sup> mice) were used. Non-littermates were also used for the platform-leaving test (see below). For the Shuttlebox and passive avoidance tests, the elevated plus-maze test, the open field test, the Morris maze, the radial-arm maze, and the Lashley maze, the strain-129 +/+ and congenic *Ntan1*<sup>-/-</sup> mice (produced, respectively, through +/+ X +/+ and *Ntan1*<sup>-/-</sup> X *Ntan1*<sup>-/-</sup> matings) were used. The same series of tests were repeated six weeks or more after the initial experiment, using both male and female mice. To determine whether the *Ntan1* genotype affected long-term memory, mice were retested in the Morris maze, Lashley maze and the Shuttlebox avoidance test at least 7 weeks after their original learning. The data were evaluated using the analysis of variance (ANOVA). Significance was set at the .05 level; all behavioral differences described in Results were at this or higher levels of statistical significance.

*Rotarod test.* The apparatus (UGO, Basile, Italy) consisted of a motor-driven rod 3 cm in diameter that carried 5 compartments, separated by walls 25 cm in height. One pair of littermates was placed on the rotating rod in each compartment. Mice were placed on the rod at either 10 or 20 rpm; the cutoff time was 2 min. Each mouse was tested on the rod 6 times, at 1 hr intervals.

*Weight retention.* Mice were suspended by the tails and made to grab a wire loop (wire diameter 2 mm; weight 40 g) 70 cm above the floor. Time elapsed before the animal released the loop was measured. Each mouse was tested 5 times a day, at 1 hr intervals, for the total of 3 days.

*Coat-hanger test.* A wire coat-hanger (wire diameter, 2 mm; length of side bars, 29 cm; length of the horizontal bar, 45 cm) suspended 75 cm above the table was used to examine motor coordination (11). A trial began by allowing the mouse to grab the middle of the horizontal bar with two front paws. The time elapsed before the animal grabbed the horizontal bar with all four paws, and the time before the two front paws reached one of the side bars were determined. Each mouse in the trial was tested

twice a day, at 6 hr intervals, for 9 days.

*Shuttlebox avoidance, passive avoidance, and elevated plus-maze tests.* For the shuttlebox and passive avoidance tests, the Gemini Avoidance system (San Diego Instruments, San Diego, CA) was used. The procedures are described in Results.

*Open field test.* For the open-field activity measurements, the Omnitech Digiscan Animal Activity Monitor was used. Each mouse was placed into a square chamber (20 x 20 cm) for one 9-min session, and the movements were tabulated along the X and Y axes, using infrared sensors.

*Morris maze.* A mouse was placed into a circular black water tub ( $d=123$  cm; temperature 19-21°C) from one of the four locations (N, S, E, W), and had to find a platform ( $d=23$  cm) submerged 1 cm below the surface of water and set 23 cm from the wall. There were no cues within the maze, so the animal had to use extra-maze spatial cues in guiding itself to the platform in repeated trials (46). An animal was given 4 trials per day (maximum of 45 sec per trial), one from each of the four locations , their order determined quasi-randomly. Each mouse was tested for 5 days, with the platform in a fixed position. On day 6, the platform was moved to the quadrant diagonally opposite its previous location, and each mouse was given 4 trials, as before. This regimen is referred to as "reversal learning". The mice were re-tested (with the platform in the original position) 7 or 8 weeks later. The animal's path was traced on an electronic digitizing tablet containing a template of the maze (22). The following parameters were recorded: total time; distance traveled; average speed; and percent of time spent in each quadrant of the maze.

*Spatial radial-arm maze.* The apparatus and testing procedure were described previously (30). Hidden escape platforms were placed at the ends of 4 of the 8 identical arms which radiated from the central area. During training the animal was released from a start arm and had to find one of the escape platforms. Once a platform was found, the arm containing that platform was blocked. The training session consisted of 4 trials in which a mouse had to locate all 4 platforms. This was followed by 11 testing sessions which differed from the training session only by removing the platform while the mouse was in its home cage between trials. Thus, on successive trials, the animal had to remember which alleys it had previously

entered and not enter those alleys again. The score was total errors obtained by summing together Working Memory Correct errors, Working Memory Incorrect errors, and Reference Memory errors (30).

*Non-spatial radial-arm maze learning.* The non-spatial swimming radial-arm maze apparatus and testing procedure were similar to the above spatial version, except for the following. Each of the arms was painted in a different black/white pattern: solid black, solid white, vertical stripes, horizontal stripes, black with white polka dots, white with black polka dots, zigzags, and the galvanized steel gray color (the start arm). The patterned arms which contained platforms were different for each mouse, but remained constant throughout the testing of a given mouse. For the training session (session 1), the entrance to the arm with the just chosen platform was blocked, and in addition, the maze was quasi-randomly rotated, so that each patterned arm now pointed toward a different place in space. As in the spatial version, the test sessions (sessions 2-12) were the same as the training session, except that while the mouse was in its home cage, after finding a platform, this platform was removed from the maze, and the arm's entrance remained open for the remainder of the session. Over 8 trials, or 2 sessions, each patterned arm pointed toward each of the 8 spatial locations in the testing room. The errors were scored as above.

*Lashley maze learning.* This maze contained cul-de-sacs that an animal had to learn to avoid, and T-choices that required the animal to learn making correct left or right turns. A water version of the maze was used, with the temperature at 19-20°C (21). All mice were given one trial per day for 5 days. An animal can learn this maze using two mutually non-exclusive strategies: by memorizing extra-maze spatial cues, or by memorizing the sequence of correct left and right turns. The measures of learning included the learning index, defined as the number of correct entries divided by total number of entries, the number of cul entries (entries into dead ends) when swimming forward (*i. e.*, toward the goal), number of forward T-choice errors, and the total number of backward errors. Retention testing was done 7 weeks later for Experiment 1, and 8 weeks later for Experiment 2.

*Platform-leaving test.* To compare the exploratory activities of two previously untested mice on the same platform, we devised the platform-leaving test. One

*Ntan1*<sup>-/-</sup> mouse of the strain 129 and one congenic +/+ mouse of the same gender (either a littermate or a non-littermate) were placed, at the same time, on the platform (16 cm x 22 cm) 2.7 cm above the surface of a laboratory bench. For each test, the platform was covered with fresh white paper. Mice were allowed to either explore or step down from the platform up until the cutoff time of 3 min. (Step-down was recorded when all four paws of a mouse were no longer in contact with platform.) The number and genotypes of mice leaving the platform first, and the number of mice not leaving by 3 min were recorded. A total of 3 trials, at 1-hr intervals, were carried out for every pair of mice tested on a given day. Each set of tests employed ~10 pairs of mice 2 to 4 months old. Four independent sets of tests, with previously untested mice, were carried out over the period of 6 months.

## RESULTS

**Construction of *Ntan1*<sup>-/-</sup> mice.** A deletion/disruption allele of the ~17 kb mouse *Ntan1* gene was constructed by replacing the *Ntan1* exons 2-5 with a neomycin-resistance cassette (Fig. 1B). Exons 2-5 encoded a 118-residue region of the 310-residue NTAN1p (Nt<sup>N</sup>-amidase) that was significantly conserved among the NTAN1p proteins of different species, including the plant *Arabidopsis thaliana*, the fly *Drosophila melanogaster*, and mammals (data not shown). Of the ~1,000 embryonic stem (ES) cell clones (strain 129/CJ7) resistant to both G418 and FIAU, 33 clones contained the expected mutation, as verified by PCR and Southern analyses (data not shown). Twelve of these correctly targeted ES cell clones were used to generate male chimeras, and in six of them the *Ntan1*<sup>-/-</sup> allele was transmitted through germ line. Male chimeras were mated with either 129/SvEv or C57BL/6 females, yielding *Ntan1*<sup>+/+</sup> heterozygotes. Intercrosses of heterozygous mice produced *Ntan1*<sup>-/-</sup> progeny at the expected frequency of approximately 1 in 4 (Figure 1C, D), indicating that the absence of NTAN1p did not increase embryonic lethality. The behavioral tests described below utilized exclusively *Ntan1*<sup>-/-</sup> mice produced through matings in the strain-129 background.

**Expression of *Ntan1* and intracellular localization of NTAN1p.** Several regions of *Ntan1* transcripts were analyzed using RT-PCR with RNA from wildtype and

*Ntan1*<sup>-/-</sup> EF cells. The level of *Ntan1*-derived RNA in the *Ntan1*<sup>-/-</sup> EF cells was below the RT-PCR detection threshold (Fig. 1E and data not shown). Northern hybridization analyses, using probes encompassing either exons 2-5 (the deleted region) or exons 1-5 detected the 1.4 and 1.6 kb *Ntan1* mRNAs, respectively, in the brains and testes of +/+ mice; the +/+ testes also contained the *Ntan1*-derived RNAs of 1.1 and 4 kb (Fig. 2A). No *Ntan1* transcripts were detected in the brains and testes of *Ntan1*<sup>-/-</sup> mice with these probes (Fig. 2Ab, d).

Similar analyses but using probes encompassing exons 1-10 or 2-10 (except for the IL2-similarity region of exon 10 (24)) detected at most trace amounts of *Ntan1*-derived transcripts in the brains of *Ntan1*<sup>-/-</sup> mice (Fig. 2Aa, c). However, in the testes of *Ntan1*<sup>-/-</sup> mice, the same probes detected 1.1 and 4 kb *Ntan1*-derived, testis-specific transcripts, but not the 1.6 kb transcript (Fig. 2Aa, c). Further mapping, using the probes shown in Figure 2C, indicated that the 1.1 and 4 kb transcripts hybridized to *Ntan1* exons 6-10 and 9-10, respectively (Fig. 2Ae, f). These transcripts were detected with antisense but not with sense probes (data not shown), indicating the direction of transcription identical to that for the full-length *Ntan1* mRNA. Thus, the testes but not the brains of *Ntan1*<sup>-/-</sup> mice contained a set of *Ntan1*-derived transcripts that lacked the 5'-half of the *Ntan1* ORF (including the region deleted in *Ntan1*<sup>-/-</sup> mice) and encompassed its 3' half. Apparently the same transcripts were present in the +/+ testes (Fig. 2Ae, f). Whether these testis-specific transcripts are physiologically relevant remains to be determined.

We also examined, using Northern analysis, whether the expression of other components of the N-end rule pathway was altered in the *Ntan1*<sup>-/-</sup> mice. No changes in the expression of either *Ubr1* (encoding the E3 of the N-end rule pathway) or *Ate1* (encoding R-transferase (34); see Introduction) were detected in the brains and testes of *Ntan1*<sup>-/-</sup> mice (Fig. 2B). The expression of *Ubr2* and *Ubr3* (homologs of *Ubr1*; Y. T. Kwon and A. Varshavsky, unpublished data) and of *mHR6B*, encoding the E2<sub>14K</sub> Ub-conjugating (E2) enzyme, was also unchanged in the *Ntan1*<sup>-/-</sup> mice (data not shown). In addition, no alterations in the ratio of *Ate1-1* mRNA to *Ate1-2* mRNA, which encode two splicing-derived forms of the mouse R-transferase (34), was

detected in the brains, testes, and EF cells of *Ntan1*<sup>-/-</sup> mice (data not shown).

*Ntan1* was expressed in most if not all tissues of adult mice (24). Whole-mount *in situ* hybridization was used to examine the expression of *Ntan1* in +/+ mouse embryos. In e9.75 (9.75 days old) embryos, *Ntan1* was most strongly expressed in the branchial arches, the tail bud and the forelimb buds; the same probe detected no signal in the *Ntan1*<sup>-/-</sup> embryos (Fig. 3A). In e10.5 and e11.5 embryos, the expression of *Ntan1* became high in the hindlimb buds as well; this expression pattern was indistinguishable from that of *Ubr1* (Fig. 3A), consistent with the assignment of NTAN1p and UBR1p to the same pathway.

To determine the intracellular location of NTAN1p, it was expressed in mouse 3T3 cells as a fusion to the N-terminus of Green Fluorescent Protein (GFP). Whereas the free 26 kDa GFP was distributed uniformly throughout the cell, the 65 kDa NTAN1p-GFP was enriched in the nucleus, was also present in the nucleus-proximal cytoplasm, but was apparently much less abundant at the cells' periphery (Fig. 3B and data not shown).

***Ntan1*<sup>-/-</sup> mice lack Nt<sup>N</sup>-amidase and the asparagine-specific branch of the N-end rule pathway.** To determine whether *Ntan1*<sup>-/-</sup> EF cells lacked the Nt<sup>N</sup>-amidase activity, we employed a previously described assay (24). The <sup>35</sup>S-labeled X-DHFR test proteins (X=Asn, Gln, or Asp) were incubated with a whole-cell extract and thereafter fractionated by isoelectrofocusing (IEF), which separated Asn-DHFR and Gln-DHFR from Asp-DHFR and Glu-DHFR, respectively. Incubation of Asn-DHFR with the extract from +/+ EF cells shifted the pI of this protein to that of Asp-DHFR. In contrast, the pI of Asn-DHFR remained unchanged after an otherwise identical incubation with the same amount of extract from *Ntan1*<sup>-/-</sup> EF cells (Fig. 4A). The pI of Gln-DHFR shifted to that of Glu-DHFR after incubations with either +/+ or *Ntan1*<sup>-/-</sup> extracts (Fig. 4A). Thus, *Ntan1*<sup>-/-</sup> EF cells lacked the Nt<sup>N</sup>-amidase activity, but contained wildtype amounts of the Gln-specific Nt<sup>Q</sup>-amidase. The latter result indicated that the putative mammalian Nt<sup>Q</sup>-amidase (which remains to be isolated and cloned) is encoded by a gene distinct from *Ntan1*.

To examine the *in vivo* degradation of N-end rule substrates in the absence of

Nt<sup>N</sup>-amidase, the immortalized +/+ and *Ntan1*<sup>-/-</sup> EF cell lines were transiently transfected with plasmids that expressed X-nsP4 $\beta$ gal test proteins (X=Met, Asn, Gln, or Arg). An X-nsP4 $\beta$ gal was expressed as part of a fusion containing the ha-tagged reference protein dihydrofolate reductase (DHFR). In this previously developed UPR (Ub/protein/reference) technique (40, 60), a DHFR-ha-UbR48-X-nsP4 $\beta$ gal fusion is cotranslationally cleaved by DUBs at the Ub-X-nsP4 $\beta$ gal junction, yielding the long-lived DHFR-ha-UbR48 reference protein and the test protein X-nsP4 $\beta$ gal. In the UPR technique, the reference protein serves as an internal control for the levels of expression, immunoprecipitation yields, sample volumes, and other sources of sample-to-sample variation, thereby increasing the accuracy of pulse-chase assays (40, 60).

The previously introduced term ID<sup>x</sup>, initial decay, *i. e.*, the extent of degradation of a protein *during* the pulse of  $x$  minutes (40), was used to describe the *in vivo* decay curves of test proteins. The ID<sup>x</sup> term is superfluous in the case of a strictly first-order decay, which is defined by a single half-life. However, the *in vivo* degradation of most proteins deviates from first-order kinetics. Specifically, the rate of degradation of short-lived proteins can be much higher *during* the pulse, in part because a newly labeled (either nascent or just-completed) polypeptide is conformationally immature, and may, consequently, be targeted for degradation more efficiently than its mature counterpart. This enhanced early degradation, previously termed the “zero-point” effect (5), is described by the parameter ID<sup>x</sup> (40). It was found that a large fraction of the zero-point effect results from the cotranslational degradation of nascent (being synthesized) polypeptides, which never reach their mature size before their destruction by processive proteolysis (G. Turner and A. Varshavsky, unpublished data).

Asn-nsP4 $\beta$ gal was metabolically unstable in +/+ EF cells: its ID<sup>10</sup> (percent degradation *during* the 10-min pulse) was ~45%, in comparison to ID<sup>10</sup> of the long-lived Met-nsP4 $\beta$ gal, taken as 100% (Fig. 4B, D). However, the same Asn-nsP4 $\beta$ gal was completely (and reproducibly) stabilized in the *Ntan1*<sup>-/-</sup> EF cells, its pulse-chase pattern becoming indistinguishable from the normally long-lived Met-nsP4 $\beta$ gal. In

contrast, both Gln-nsP4 $\beta$ gal and Arg-nsP4 $\beta$ gal were comparably short-lived in both +/+ and *Ntan1*<sup>-/-</sup> cells (Fig. 4B-D). Taken together with the measurements of Nt<sup>N</sup>-amidase enzymatic activity (Fig. 4A), these data indicated that the *Ntan1*<sup>-/-</sup> EF cells, in contrast to congenic wildtype cells, lacked the Asn-specific branch of the N-end rule pathway but retained the rest of this pathway.

**Motor coordination and spontaneous activity of *Ntan1*<sup>-/-</sup> mice.** The *Ntan1*<sup>-/-</sup> mice were fertile, apparently healthy, of similar size and weight as +/+ littermates, and cared for their offspring. These mice oriented to sound; their limb movements and behavior appeared to be indistinguishable from those of congenic +/+ mice. Histological examination of *Ntan1*<sup>-/-</sup> tissues (small intestine, liver, pancreas, adrenal gland, thyroid gland, kidney, ovary, testis, heart, spleen, thymus, skeletal muscle, brain, and sciatic nerve) did not detect abnormalities. No significant differences were observed between +/+ and *Ntan1*<sup>-/-</sup> thymocytes in their ability to undergo apoptosis in response to either radiation or dexamethasone, suggesting that *Ntan1*<sup>-/-</sup> mice were not impaired in apoptotic responses. Thus far, the only non-behavioral differences between +/+ and congenic *Ntan1*<sup>-/-</sup> mice were a weaker mitogenic response of *Ntan1*<sup>-/-</sup> splenocytes and thymocytes to phytohemagglutinin, and hypersensitivity of *Ntan1*<sup>-/-</sup> mice to bacterial lipopolysaccharide (preliminary data; not shown). In addition, over the last 1.5 years the frequency of natural death among *Ntan1*<sup>-/-</sup> mice was ~50% higher than among congenic +/+ mice.

Motor coordination of *Ntan1*<sup>-/-</sup> and congenic +/+ mice was compared using the Rotarod apparatus. No difference in the ability of mice to stay on a rotating horizontal rod was observed between the *Ntan1*<sup>-/-</sup> and +/+ strains (Fig. 5A). Comparisons of the physical strength (the weight retention test), of both coordination and strength (the coat-hanger test), and of the walking patterns (the hindpaw footprint test) also did not distinguish between *Ntan1*<sup>-/-</sup> and +/+ mice (Figure 5B, C and data not shown). These and related observations indicated that *Ntan1*<sup>-/-</sup> mice were not impaired in either motor coordination, physical strength, or the learning ability required for their improved performance in repeated trials (Figure A, C).

The locomotor activity of *Ntan1*<sup>-/-</sup> mice was assayed in the open-field test. By several criteria, *Ntan1*<sup>-/-</sup> mice exhibited significantly lower spontaneous activity than the congenic wildtype mice (Fig. 5F). Specifically, *Ntan1*<sup>-/-</sup> mice traveled significantly less distance (Fig. 5Fa), spent significantly more time near the field's center (Fig. 5Fb), and also, consequently, less time within 1 cm from the walls (Fig. 5Fc). These findings will be considered in conjunction with data from the platform-leaving test (see Fig. 6 and Discussion).

**Active and passive avoidance learning.** The shuttlebox avoidance test compared shock-motivated learning in +/+ and congenic *Ntan1*<sup>-/-</sup> mice. In this active avoidance learning procedure, a light-conditioned stimulus (CS) precedes, by 5 sec, the unconditioned stimulus (UCS) of electric shock, which remains on for 20 sec. The animal can (i) cross to the safe compartment within the 5-second CS interval, thus *avoiding* shock; (ii) cross over during the 20-second UCS interval, to *escape* the ongoing shock; or (iii) not cross over and receive the full 20 sec of shock (the latter outcome is scored as the *null* response). Neither +/+ nor *Ntan1*<sup>-/-</sup> mice showed any evidence of learning over the 5 days of testing. However, there were performance differences in the distribution of their responses. The two groups averaged only 4 avoidance responses per day, but the +/+ mice exhibited more escape responses and therefore fewer null responses than the *Ntan1*<sup>-/-</sup> mice (Fig. 5D). Neither group showed evidence of learning during the 5 days of retention testing, but the number of null responses by *Ntan1*<sup>-/-</sup> mice decreased significantly in comparison to their performance during learning (data not shown), indicating that memory consolidation had occurred in this group between the original testing and retesting 7-10 weeks later. Strain-129 mouse substrains are known to perform poorly in active avoidance tests (7, 53).

In the above shuttlebox test, an electric shock is used to compel the animal to make an active response. By contrast, in passive learning, the shock is used to compel the animal to inhibit its behavior. The mouse was placed into a chamber. Ten seconds later, the chamber was illuminated and a guillotine door opened, exposing a dark chamber. The natural behavior is to move into the dark. When this occurred,

the guillotine door closed and the animal received 0.4 mA of footshock for 2 seconds. The time it took the mouse to enter the dark chamber (until the cutoff time of 400 sec) was recorded. The mouse was put back in its home cage; 24 hr later, the mouse was placed back into the original chamber, and all stimulus conditions were repeated except that no shock occurred when the mouse entered the dark chamber. The usual finding is that the mouse takes longer to enter the dark chamber on the second day because the prior shock experience inhibits its behavior. Figure 5E shows the time before entry into the dark chamber ("latency time") on both days for the two genotypes. On the first day, the *Ntan1*<sup>-/-</sup> mice took significantly longer to enter the dark chamber than the +/+ animals. On the second day, the entry into the dark chamber by the *Ntan1*<sup>-/-</sup> mice was further delayed (indicating learning), while the entry by the +/+ mice was delayed even more, resulting in similar retention times for either genotype (Fig. 5E).

The longer retention time of the *Ntan1*<sup>-/-</sup> mice on the first day of passive learning, as well as their lower activity in the open-field test suggested that +/+ and congenic *Ntan1*<sup>-/-</sup> mice might differ in their anxiety levels. To investigate this, they were tested in the elevated plus maze, which consists of two open arms and two closed arms, which radiate from a center platform. The open arms are more anxiety-inducing, in that the animals with higher anxiety tend to decrease exploration of the open arms (42). Although the *Ntan1*<sup>-/-</sup> mice tended to explore the open arms less frequently than congenic +/+ mice, the observed difference was not statistically significant (data not shown).

**A socially conditioned exploratory phenotype of *Ntan1*<sup>-/-</sup> mice.** To address the finding of lower activity in the *Ntan1*<sup>-/-</sup> mice (Fig. 5F) in a different way, we devised a simple test in which the exploratory activity was assessed in the context of a social interaction. In this "platform-leaving" test, two previously untested mice, *Ntan1*<sup>-/-</sup> and a congenic +/+, are placed, close together and at the same time, onto the center of a 16 x 22 cm platform 2.7 cm in height. Thereafter one records the following variables: the time it takes each mouse to leave the platform, until the cutoff time of 3 min; the number and genotype of mice leaving the platform first (in independent trials with

pairs of mice); and the number and genotypes of mice not leaving the platform by the cutoff time (see Materials and Methods). In contrast to the standard open-field test, a mouse in the platform-leaving test sees and smells the second mouse on the same platform, and is otherwise influenced by the second mouse in the course of their peregrinations over the platform. Thus the proclivity of a mouse to explore its environment is modulated, in the platform-leaving test, by interactions with another mouse.

A striking result of this test was the observation that an *Ntan1*<sup>-/-</sup> mouse paired with its congenic <sup>+/+</sup> littermate tended to stay longer on the platform, and almost never left it by the cutoff time, in contrast to the <sup>+/+</sup> mouse (Fig. 6A). For example, in a trial with 44 pairs of previously untested *Ntan1*<sup>-/-</sup> and <sup>+/+</sup> mice, only 1 *Ntan1*<sup>-/-</sup> mouse left the platform by the cutoff time, whereas 20 <sup>+/+</sup> mice left it by that time (Fig. 6Ac). Repeated trials, at 1-hour intervals, in the platform-leaving test progressively decreased this initial difference between the paired *Ntan1*<sup>-/-</sup> and <sup>+/+</sup> mice (Fig. 6A). Note that this decrease resulted from both an *increase* in the exploratory activity of *Ntan1*<sup>-/-</sup> mice (in successive trials) and a *decrease* in the exploratory activity of their paired <sup>+/+</sup> littermates (e. g., Fig. 6Ac). Thus, the strongly different initial responses of paired *Ntan1*<sup>-/-</sup> and <sup>+/+</sup> mice to the novel environment (of which the other mouse was a part) became attenuated in subsequent trials (Fig. 6A), and nearly disappeared by the 7th trial (data not shown).

We then asked whether the observed difference between *Ntan1*<sup>-/-</sup> and <sup>+/+</sup> mice was retained if the same test was performed with congenic *non-littermates*. Strikingly, if the test of Figure 6A was carried out with wildtype and mutant non-littermates, the results were reproducibly different from those with littermates (Fig. 6B). Specifically, in this case *Ntan1*<sup>-/-</sup> mice did not stay longer on the platform: they left it, on average, more frequently than <sup>+/+</sup> mice, a pattern opposite to that seen with littermates (Fig. 6Bb; compare with Fig. 6Ab). In addition, the striking difference between *Ntan1*<sup>-/-</sup> and <sup>+/+</sup> littermates in their frequency of not leaving the platform by the cutoff time was no longer seen in the otherwise identical tests with *non-littermates* (Fig. 6Ac; compare with Fig. 6Bc).

In the control experiments, multiple pairs of littermates of the *same* genotype (either both +/+ or both *Ntan1*<sup>-/-</sup>) were subjected to this test. No statistically significant differences between the platform-leaving patterns of the two mice in a pair were observed (data not shown). The results described in Figure 6 were consistently reproduced in 4 independent sets of tests, carried out over 6 months with previously untested mice (at least 10 pairs of mice per test). This socially conditioned exploratory phenotype of *Ntan1*<sup>-/-</sup> mice (Fig. 6) has not been previously described, to our knowledge, with other mouse strains.

**Spatial learning and memory.** To compare spatial learning of +/+ and congenic *Ntan1*<sup>-/-</sup> mice, we tested them on the hidden platform Morris maze (where a mouse had to find a platform submerged in water pool), using the "reversal learning" regimen (see Materials and Methods) and re-testing 7 or 8 weeks later. Both +/+ and *Ntan1*<sup>-/-</sup> mice learned to find the platform, and did not differ in either the time they took to do it or the distance traveled; they also did not differ on these measures in the subsequent reversal-learning test, when the platform was moved to the diagonally opposite quadrant (data not shown). However, 7-8 weeks later, on the first day of retention testing, the *Ntan1*<sup>-/-</sup> mice were inferior to congenic +/+ mice, in that they took significantly longer time to find the platform, and traveled a significantly longer distance before reaching it (data not shown), suggesting that the *Ntan1*<sup>-/-</sup> mice have a less effective spatial memory system.

Another way to study spatial learning employs the water version of the radial arm maze, where 4 of the 8 arms contain escape platforms (30). There are no intra-maze cues available to the mouse, and it must use extra-maze spatial information involving reference and working memory to remember which alleys to enter and which to avoid. The mice were given one training day and were tested for 11 consecutive days, receiving 4 trials each day. The measure used was the total number of errors. *Ntan1*<sup>-/-</sup> mice made significantly more errors in the acquisition phase (sessions 2-7) than the congenic +/+ mice (Fig. 7A). In the asymptotic phase (sessions 8-12), there were no significant differences between the two strains (Fig. 7A).

The 4 trials within each daily session impose an increasing demand on memory. Thus, on the first trial, there are platforms in 4 of the 8 alleys, and entering any of these terminates the trial. On the second trial, 3 of the alleys still contain platforms; the mouse has to remember which alley it had entered on the prior trial, and not enter that alley again. Finally, on the fourth trial, the mouse has to remember which three alleys it had chosen earlier, so that it can find the one remaining alley containing a platform. Thus, by studying the error rate from trial to trial, summed over the sessions, one can compare the ability of mice to cope with an increasing memory load. Figure 7C shows the error curves summed over the acquisition phase. As expected, errors increased as the memory load increased. However, the rate of increase was significantly greater for the *Ntan1*<sup>-/-</sup> mice, especially on Trial 4 (Fig. 7C).

Unlike previously tested mouse strains (30), both +/+ and congenic *Ntan1*<sup>-/-</sup> mice exhibited poor learning over the 11 days of testing, consistent with inferior performance of strain 129-derived mice in several other learning regimens (7, 53). Thus, the significant effects described above (Fig. 7A, C) represent largely differences in performance, but say little about the spatial learning ability of these mice. To determine whether the observed effects were related to spatial competence, the identical experiment was carried out using the same radial-arm maze, except that a *non-spatial* version was devised, by having different visual patterns in each alley and rotating the maze between trials so that the extra-maze cues were rendered meaningless. In 4 of the arms, escape platforms were placed, and the mice had to learn to associate a specific visual configuration (e.g., black dots on a white background, or horizontal black and white stripes) with the presence or absence of a platform (29). In contrast to the results with the spatial radial-arm maze, no significant differences between the two genotypes were observed in the non-spatial radial-arm maze, either in regard to the overall number of errors (Fig. 7B) or in regard to the error rate as a function of memory load (data not shown). Here, too, both +/+ and congenic *Ntan1*<sup>-/-</sup> mice did not exhibit significant learning over the 11 days of testing. Thus, no conclusions can be drawn about spatial or non-spatial learning. At

the same time, it was possible to conclude that *Ntan1*<sup>-/-</sup> mice were less effective in dealing with spatial information than the congenic +/+ mice.

Another test of spatial learning and memory was the Lashley III maze. This maze contains cul-de-sacs that an animal has to learn to avoid, and T-choices that require the animal to make the correct left or right turn. Both +/+ and *Ntan1*<sup>-/-</sup> mice were able to learn the maze, but *Ntan1*<sup>-/-</sup> mice were significantly less effective, as measured by the learning index, T-forward errors (making a wrong T-choice), and the total number of backward errors (Fig. 7D). When retested 7 or 8 weeks later, *Ntan1*<sup>-/-</sup> mice had lower scores on the first trial on all three measures, showing inferior long-term retention (data not shown). These results were similar to the findings with the Morris maze (see above), except that there was a significant difference in the original learning of the Lashley maze favoring the +/+ mice (Fig. 7D). This difference may stem from the fact that the Lashley maze can be learned using intra-maze cues to memorize the pathway of left and right turns, in addition to the use of extra-maze cues, which are present in both the Lashley and Morris maze tests. The results of the two radial-arm studies (Fig. 7A-C) indicated that the *Ntan1*<sup>-/-</sup> mice used spatial information less effectively than the +/+ mice. In summary, the results of spatial memory tests (Fig. 7) strongly suggested that mouse Nt<sup>N</sup>-amidase contributes to the processing of spatial information and to long-term retention of spatial learning.

## Discussion

Homologs of the 310-residue mammalian Nt<sup>N</sup>-amidase, encoded by the *Ntan1* gene (24, 59), are present in other vertebrates, in arthropods such as *D. melanogaster*, and in plants such as *A. thaliana*, but are apparently absent from the nematode *C. elegans* (data not shown). The deamidation of N-terminal (and apparently only N-terminal) Asn residue in proteins or short peptides is the only known enzymatic activity of Nt<sup>N</sup>-amidase. This enzyme was operationally defined as a component of the N-end rule pathway, because it converts, *in vivo*, the N-terminal Asn of engineered N-end rule substrates into N-terminal Asp, which is then conjugated by R-transferase to Arg, a primary destabilizing residue (34, 63). Physiological substrates of Nt<sup>N</sup>-amidase remain to be identified. Physiological substrates are also unknown

for the fungal Nt-amidase, which can deamidate either N-terminal Asn or N-terminal Gln, and lacks significant sequence similarities to metazoan Nt<sup>N</sup>-amidase (6). In the present work, mouse Nt<sup>N</sup>-amidase was studied using several approaches, including targeted mutagenesis. We report the following results:

(i) Both copies of the mouse *Ntan1* gene, which encodes the Asn-specific N-terminal amidohydrolase (Nt<sup>N</sup>-amidase) were replaced by a deletion/disruption allele. The resulting *Ntan1*<sup>-/-</sup> mice were viable and fertile. Their behavioral phenotypes are described below.

(ii) Embryonic fibroblast (EF) cells from *Ntan1*<sup>-/-</sup> mice lacked the Nt<sup>N</sup>-amidase activity, in contrast to congenic +/+ EF cells. In addition, among the normally short-lived substrates of the N-end rule pathway, only those bearing N-terminal Asn, a tertiary destabilizing residue (see Introduction), became long-lived in the *Ntan1*<sup>-/-</sup> EF cells. Thus, at least these cells, and, by inference, the *Ntan1*<sup>-/-</sup> mice, lacked the asparagine branch of the N-end rule pathway.

(iii) The *Ntan1*<sup>-/-</sup> EF cells contained wildtype levels of Gln-specific Nt-amidase (Nt<sup>Q</sup>-amidase), which mediates the activity of N-terminal Gln, another tertiary destabilizing residue in the N-end rule. Thus, Nt<sup>N</sup>-amidase and Nt<sup>Q</sup>-amidase (the latter enzyme remains to be isolated and characterized) are encoded by different genes.

(iv) The brains and testes of +/+ mice contained, respectively, the 1.4 and 1.6 kb *Ntan1* mRNAs. In addition, the testes, but not the brains of +/+ mice also contained the *Ntan1*-derived RNAs of 1.1 and 4 kb. The brains and testes of *Ntan1*<sup>-/-</sup> mice lacked, respectively, the 1.4 and 1.6 kb *Ntan1* mRNAs. However, the testes of *Ntan1*<sup>-/-</sup> mice retained the testis-specific 1.1 and 4 kb *Ntan1*-derived RNAs, which were found to lack the 5'-half of the *Ntan1* ORF (including the region deleted in *Ntan1*<sup>-/-</sup> mice), and to encompass the 3' half of *Ntan1* ORF. Whether these testis-specific transcripts are physiologically relevant in +/+ mice remains to be determined.

(v) Deletion/disruption of *Ntan1* in the *Ntan1*<sup>-/-</sup> mice did not significantly change the levels of mRNAs encoding other components of the N-end rule pathway,

such as ATE1p (R-transferase); UBR1p (E3 $\alpha$ , the pathway's E3 component); UBR2p (a homolog of UBR1p); and mHR6Bp (the E2<sub>14K</sub> Ub-conjugating (E2) enzyme).

(vi) *Ntan1*, which is expressed, at different levels, in most if not all tissues of adult mice (24), was found, through *in situ* hybridization, to be particularly strongly expressed in the branchial arches, the tail bud and the limb buds of 10 days old embryos. This expression pattern was indistinguishable from that of *Ubr1* (36), consistent with the assignment of NTAN1p (Nt<sup>N</sup>-amidase) and UBR1p (E3 $\alpha$ ) to the same pathway.

(vii) An NTAN1p-GFP fusion was enriched in the nucleus of transfected mouse 3T3 cells, in contrast to free GFP, which was uniformly distributed. NTAN1p-GFP was also present in the nucleus-proximal cytoplasm, but was apparently much less abundant at the cells' periphery. This localization pattern of the mammalian Nt<sup>N</sup>-amidase is quite different from that of the *S. cerevisiae* *NTA1*-encoded Nt-amidase, most of which is located in the yeast mitochondria (H. R. Wang and A. Varshavsky, unpublished data). In contrast to the mammalian Nt<sup>N</sup>-amidase, the yeast Nt-amidase, which has no significant sequence similarities to the metazoan enzyme, can deamidate either N-terminal Asn or N-terminal Gln (6).

(viii) While the abnormal phenotypes of *Ntan1*<sup>-/-</sup> mice are apparently not confined to behavioral/cognitive alterations (see Results), it is the latter that have been analyzed, thus far, in some detail. Briefly, the *Ntan1*<sup>-/-</sup> mice were indistinguishable from congenic +/+ mice in motor coordination and general physical performance, but had reduced spontaneous activity and less effective spatial memory. Remarkably, the exploratory behavior of *Ntan1*<sup>-/-</sup> mice was found to be particularly different from that of congenic +/+ mice in the presence of social interactions. If a previously untested *Ntan1*<sup>-/-</sup> mouse was placed on a small, slightly elevated platform with its +/+ littermate (which it grew up with in the same cage), the *Ntan1*<sup>-/-</sup> mouse left the platform much more slowly than the +/+ mouse. In contrast, when an *Ntan1*<sup>-/-</sup> mouse was subjected to this test in the presence of a congenic non-littermate +/+ mouse, it tended to leave the platform *faster* than the

$+/+$  mouse, in spite of the fact that  $Ntan1^{-/-}$  mice exhibited diminished exploratory activity when tested alone (Fig. 6).

Our working hypothesis is that  $Ntan1^{-/-}$  mice are socially recessive, in relation to congenic  $+/+$  mice, and in addition have a lower spontaneous exploratory activity. In a relatively unstressful setting, such as when it is placed on the platform side-by-side with a (familiar) littermate, the behavior of  $Ntan1^{-/-}$  mice is governed largely by their inherently lower exploratory activity, yielding the observation that these mice leave the platform much more slowly than their  $+/+$  littermates (Fig. 6A). By contrast, in an otherwise identical test with a  $+/+$  non-littermate, the postulated social recessiveness of  $Ntan1^{-/-}$  mice becomes their behavior-governing trait. As a result, they leave the platform faster than the  $+/+$  non-littermates, to reduce the proximity-induced anxiety (Fig. 6B).

The socially conditioned exploratory phenotype of  $Ntan1^{-/-}$  mice, identified through the platform-leaving test (Fig. 6), has not been previously described, to our knowledge, with other mouse strains. It is difficult to compare this finding with behavioral studies of other mouse mutants, because most of the earlier analyses of exploratory behavior employed solitary mice. An example of differential effect of stress on a specific mutant is provided by the analysis of mice lacking the serotonin receptor 1A. These mice exhibited decreased exploratory activity and increased fear of aversive environments (in comparison to  $+/+$  mice), as measured by the open field test and the elevated plus-maze test, respectively. However, the mutant mice were indistinguishable from  $+/+$  littermates in a less stressful setting such as their home cages (50).

A parsimonious molecular interpretation of our results (Figs. 5-7) is that a normally short-lived regulatory protein(s) that is targeted for degradation by the N-end rule pathway through the protein's N-terminal Asn becomes long-lived in  $Ntan1^{-/-}$  mice. The resulting increase in the steady-state concentration of this protein(s) alters the functioning of relevant neural networks in the adult brain or, non-alternatively, changes these networks in the course of their establishment during development. That the inactivation of a Ub-dependent proteolytic pathway could affect cognitive functions is illustrated, for example, by the identification of *UBE3A* as

a human gene encoding an E3 protein of the Ub system called E6AP. This gene is mutated in the Angelman's syndrome, the symptoms of which include motor dysfunction and mental retardation (33, 45). Mice lacking *UBE3A* exhibit deficits in contextual learning and long-term potentiation (31).

The absence of severe impairments in the *Ntan1*<sup>-/-</sup> mice should make them particularly suitable as vehicles for the approaches to conditional mutagenesis that utilize conditional destabilization of a protein of interest. One version of this approach could utilize a transgenic mouse strain that lacks endogenous *Ntan1* and expresses this gene from one of the previously constructed promoters whose activity can be controlled by small molecules such as tetracycline or ecdysone (9). If a protein of interest is modified, through knock-in mutagenesis and the Ub fusion technique (63), to bear an Asn-containing N-degron, the resulting protein would be long-lived in the absence of the *Ntan1*-encoded Nt<sup>N</sup>-amidase but short-lived, and therefore scarce, in the presence of Nt<sup>N</sup>-amidase. In this method, the steady-state level, and hence the activity of a protein of interest, is controlled through *Ntan1*-dependent, regulated changes of the protein's metabolic stability. One advantage of this strategy is that it does not require alterations of a promoter that expresses a gene of interest, thereby avoiding potential perturbations of the promoter regulation during development and differentiation. This approach may be combined with the existing technologies for conditional mutagenesis of the mouse (13, 55), enhancing them through regulated degradation of a protein of interest.

Inasmuch as behavior is an emergent property of highly complex neural networks interacting with the musculoskeletal apparatus, a behavioral alteration that is caused by the absence of a specific protein from the brain is difficult to understand in terms of higher-order neural events even in the case of a protein (e. g., a neurotransmitter-gated ion channel) whose "local" molecular function in the individual neurons is clear. The difficulty is further increased in the case of a protein such as Nt<sup>N</sup>-amidase, whose physiological substrates remain to be identified. At the same time, the absence of *a priori* reasons to suspect a specific role for Nt<sup>N</sup>-amidase in the brain's functions, and the robustness of behavioral differences between the *Ntan1*<sup>-/-</sup> and congenic +/+ mice (Figs. 5-7) make these findings particularly

intriguing. Further advances in this inquiry will require, at minimum, the identification of physiological substrates of Nt<sup>N</sup>-amidase. The availability of *Ntan1*<sup>-/-</sup> mice and cell lines derived from them should facilitate the discovery of these substrates.

### Acknowledgments

We are grateful to members of the Caltech Transgenic Facility, especially S. Pease, B. Kennedy, and A. Granados, for the care of mice and expert technical help. We thank N. Barteneva for assistance with some of the early experiments, S. Offermanns for advice and help with preparation of embryonic fibroblasts, and R. C. Mulligan and R. Jaenisch for their gifts of plasmids. This work was supported by grants from the National Institutes of Health to A. V. (GM31530 and DK39520).

### Footnote

<sup>†</sup>Throughout the text, the names of mouse genes are in italics, with the first letter upper-case. The names of human and *Saccharomyces cerevisiae* genes are also in italics, all upper-case. If human and mouse genes are named in the same sentence, the mouse gene notation is used. The names of *S. cerevisiae* proteins are Roman, with the first letter upper-case and an extra lower-case "p" at the end. The names of mouse and human proteins are the same, except that all letters but the last "p" are upper-case. The latter usage is a modification of the existing convention (58), to facilitate simultaneous discussions of yeast, mouse and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.)

**REFERENCES**

1. Alagramam, K., F. Naider, and J. M. Becker. 1995. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. Mol. Microbiol. 15:225-234.
2. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.). 1998. Current Protocols in Molecular Biology. Wiley-Interscience, New York.
3. Bachmair, A., D. Finley, and A. Varshavsky. 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science 234:179-186.
4. Bachmair, A., and A. Varshavsky. 1989. The degradation signal in a short-lived protein. Cell 56:1019-1032.
5. Baker, R. T., and A. Varshavsky. 1991. Inhibition of the N-end rule pathway in living cells. Proc. Natl. Acad. Sci. USA 87:2374-2378.
6. Baker, R. T., and A. Varshavsky. 1995. Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. J. Biol. Chem. 270:12065-12074.
7. Balogh, S. A., C. S. McDowell, A. Stavnezer, and V. H. Denenberg. 1999. A behavioral and neuroanatomical assessment of an inbred substrain of 129 mice, with behavioral comparisons to C57BL/6J mice. Brain Res. 836:38-48.
8. Balzi, E., M. Choder, W. Chen, A. Varshavsky, and A. Goffeau. 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. J. Biol. Chem. 265:7464-7471.
9. Baron, U., D. Schnappinger, V. Helbl, M. Gossen, W. Hillen, and H. Bujard. 1999. Generation of conditional mutants in higher eukaryotes by switching between the expression of two genes. Proc. Natl. Acad. Sci. USA 96:1013-1018.
10. Baumeister, W., J. Walz, F. Zühl, and E. Seemüller. 1998. The proteasome: paradigm of a self-compartmentalizing protease. Cell 92:367-380.
11. Beaudin, S., and R. Lalonde. 1997. The effects of pentobarbital on spatial learning, motor coordination, and exploration. Pharmacol. Biochem. Behav. 57:111-114.
12. Byrd, C., G. C. Turner, and A. Varshavsky. 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor.

- EMBO J. 17:269-277.
13. **Capecci, M. R.** 1989. Altering the genome by homologous recombination. Science 244:1288-1292.
  14. **Chang, S. E., J. Keen, E. B. Lane, and J. Taylor-Papadimitriou.** 1982. Establishment and characterization of SV40-transformed human breast epithelial cell lines. Cancer Res. 42:2040-2053.
  15. **Chau, V., J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky.** 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243:1576-1583.
  16. **Conlon, R. A., and J. Rossant.** 1992. Exogenous retinoic acid rapidly induces anterior ectopic expression of murine Hox-2 genes *in vivo*. Development 116:357-368.
  17. **Coux, O., K. Tanaka, and A. L. Goldberg.** 1996. Structure and functions of the 20S and 26S proteasomes. Annu. Rev. Biochem. 65:801-817.
  18. **Davydov, I. V., D. Patra, and A. Varshavsky.** 1998. The N-end rule pathway in *Xenopus* egg extracts. Arch. Biochem. Biophys. 357:317-325.
  19. **deGroot, R. J., T. Rümenapf, R. J. Kuhn, and J. H. Strauss.** 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. USA 88:8967-8971.
  20. **DeMartino, G. N., and C. A. Slaughter.** 1999. The proteasome, a novel protease regulated by multiple mechanisms. J. Biol. Chem. 274:22123-22126.
  21. **Denenberg, V. H., N. Talgo, D. A. Carroll, S. Freter, and R. Deni.** 1991. A computer-aided procedure for measuring Lashley III maze performance. Physiol. Behav. 50: 857-861.
  22. **Denenberg, V. H., N. W. Talgo, N. S. Waters, and G. H. Kenner.** 1990. A computer-aided procedure for measuring swim rotation. Physiol. Behav. 47:1023-1025.
  23. **Gonda, D. K., A. Bachmair, I. Wünning, J. W. Tobias, W. S. Lane, and A. Varshavsky.** 1989. Universality and structure of the N-end rule. J. Biol. Chem. 264:16700-16712.
  24. **Grigoryev, S., A. E. Stewart, Y. T. Kwon, S. M. Arfin, R. A. Bradshaw, N. A. Jenkins, N. G. Copeland, and A. Varshavsky.** 1996. A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule

- pathway. *J. Biol. Chem.* **271**:28521-28532.
25. **Haas, A. J., and T. J. Siepman.** 1997. Pathways of ubiquitin conjugation. *FASEB J.* **11**:1257-1268.
26. **Hershko, A., and A. Ciechanover.** 1998. The ubiquitin system. *Annu. Rev. Biochem.* **76**:425-479.
27. **Hochstrasser, M.** 1996. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**:405-439.
28. **Hondermarck, H., J. Sy, R. A. Bradshaw, and S. M. Arfin.** 1992. Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. *Biochem. Biophys. Res. Comm.* **30**:280-288.
29. **Hyde, L. A., and V. H. Denenberg.** 1999. BXSB mice can learn complex pattern discrimination. *Physiol. Behav.* **66**:437-439.
30. **Hyde, L. A., B. J. Hoplight, and V. H. Denenberg.** 1998. Water version of the radial-arm maze: learning in three inbred strains of mice. *Brain Res.* **785**:236-244.
31. **Jiang, Y. H., D. Armstrong, U. Albrecht, C. M. Atkins, J. L. Noebels, G. Eichele, J. D. Sweatt, and A. L. Beaudet.** 1998. Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. *Neuron* **21**:799-811.
32. **Johnson, E. S., D. K. Gonda, and A. Varshavsky.** 1990. *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. *Nature* **346**:287-291.
33. **Kishino, T., M. Lalande, and J. Wagstaff.** 1997. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat. Genet.* **15**:70-73.
34. **Kwon, Y. T., A. S. Kashina, and A. Varshavsky.** 1999. Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. *Mol. Cell. Biol.* **19**:182-193.
35. **Kwon, Y. T., F. Lévy, and A. Varshavsky.** 1999. Bivalent inhibitor of the N-end rule pathway. *J. Biol. Chem.* **274**:18135-18139.
36. **Kwon, Y. T., Y. Reiss, V. A. Fried, A. Hershko, J. K. Yoon, D. K. Gonda, P. Sangan, N. G. Copeland, N. A. Jenkins, and A. Varshavsky.** 1998. The mouse and

- human genes encoding the recognition component of the N-end rule pathway. Proc. Natl. Acad. Sci. USA **95**:7898-7903.
37. Laney, J. D., and M. Hochstrasser. 1999. Substrate targeting in the ubiquitin system. Cell **97**:427-430.
38. Lawson, T. G., D. L. Gronros, P. E. Evans, M. C. Bastien, K. M. Michalewich, J. K. Clark, J. H. Edmonds, K. H. Graber, J. A. Werner, B. A. Lurvey, and J. M. Cate. 1999. Identification and characterization of a protein destruction signal in the encephalomyocarditis virus 3C protease. J. Biol. Chem. **274**:9871-9880.
39. Lecker, S. H., V. Solomon, S. R. Price, Y. T. Kwon, W. E. Mitch, and A. L. Goldberg. 1999. Ubiquitin conjugation by the N-end rule pathway and mRNAs for its components increase in muscles of diabetic rats. J. Clin. Invest. **104**:1411-1420.
40. Lévy, F., N. Johnsson, T. Rumenapf, and A. Varshavsky. 1996. Using ubiquitin to follow the metabolic fate of a protein. Proc. Natl. Acad. Sci. USA **93**:4907-4912.
41. Li, J., and C. Pickart. 1995. Inactivation of arginyl-tRNA protein transferase by a bifunctional arsenoxide: identification of residues proximal to arsenoxide site. Biochemistry **34**:139-147.
42. Lister, R. G. 1987. The use of a plus-maze to measure anxiety in the mouse. Psychopharmacology **92**:180-185.
43. Madura, K., and A. Varshavsky. 1994. Degradation of G $\alpha$  by the N-end rule pathway. Science **265**:1454-1458.
44. Maniatis, T. 1999. A ubiquitin ligase complex essential for the NF- $\kappa$ B, Wnt/Wingless, and Hedgehog signaling pathways. Genes Dev. **13**:505-510.
45. Matsuura, T., J. S. Sutcliffe, P. Fang, R. J. Galjaard, Y. H. Jiang, C. S. Benton, J. M. Rommens, and A. L. Beaudet. 1997. *De novo* truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat. Genet. **15**:74-77.
46. Morris, R. 1984. Development of a water-maze procedure for studying spatial learning in the rat. J. Neurosci. Meth. **11**:47-60.
47. Ota, I. M., and A. Varshavsky. 1993. A yeast protein similar to bacterial two-component regulators. Science **262**:566-569.
48. Peters, J.-M., R. W. King, and R. J. Deshaies. 1998. Cell cycle control by ubiquitin-dependent proteolysis, pp. 345-387. In J. M. Peters, J. R. Harris, and D. Finley (ed.),

- Ubiquitin and the Biology of the Cell. Plenum Press, New York, NY.
49. **Pickart, C. M.** 1997. Targeting of substrates to the 26S proteasome. *FASEB J.* **11**:1055-1066.
50. **Ramboz, S., R. Oosting, D. A. Amara, H. F. Kung, P. Blier, M. Mendelsohn, J. J. Mann, D. Brunner, and R. Hen.** 1998. Serotonin receptor 1A knockout: an animal model of anxiety-related disorder. *Proc. Natl. Acad. Sci. USA* **95**:14476-14481.
51. **Rechsteiner, M.** 1998. The 26S proteasome., pp. 147-189. *In* J. M. Peters, J. R. Harris, and D. Finley (ed.), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY.
52. **Robertson, E. J.** 1987. Embryo-derived stem cell lines., p. 71-112. *In* E. J. Robertson (ed.), *Teratocarcinomas and embryonic stem cells: A practical approach*. IRL Press, Oxford, U.K.
53. **Royce, J. R.** 1972. Avoidance conditioning in nine strains of inbred mice using optimal stimulus parameters. *Behav. Genetics* **2**:107-110.
54. **Schauber, C., L. Chen, P. Tongaonkar, I. Vega, and K. Madura.** 1998. Sequence elements that contribute to the degradation of yeast G-alpha. *Genes Cells* **3**:307-319.
55. **Schwenk, F., R. Kuhn, P. O. Angrand, K. Rajewsky, and A. F. Stewart.** 1998. Temporally and spatially regulated somatic mutagenesis in mice. *Nucl. Acids Res.* **26**:1427-1432.
56. **Sijts, A. J., I. Pilip, and E. G. Pamer.** 1997. The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class I antigen processing of bacterial proteins. *J. Biol. Chem.* **272**:19261-19268.
57. **Solomon, V., V. Baracos, P. Sarraf, and A. Goldberg.** 1998. Rates of ubiquitin conjugation increase when muscles atrophy, largely through activation of the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **95**:12602-12607.
58. **Stewart, A.** 1995. Trends in Genetics Nomenclature Guide. Elsevier Science, Ltd., Cambridge, United Kingdom.
59. **Stewart, A. E., S. M. Arfin, and R. A. Bradshaw.** 1995. The sequence of porcine protein NH<sub>2</sub>-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. *J. Biol. Chem.* **270**:25-28.

60. **Suzuki, T., and A. Varshavsky.** 1999. Degradation signals in the lysine-asparagine sequence space. *EMBO J.* **18**:6017-6026.
61. **Taban, C. H., H. Hondermarck, R. A. Bradshaw, and B. Boilly.** 1996. Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. *Experientia* **52**:865-870.
62. **Tobias, J. W., and A. Varshavsky.** 1991. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **266**:12021-12028.
63. **Varshavsky, A.** 1996. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**:12142-12149.
64. **Varshavsky, A.** 1997. The ubiquitin system. *Trends Biochem. Sci.* **22**:383-387.
65. **Wang, Y. M., and N. A. Ingoglia.** 1997. N-terminal arginylation of sciatic nerve and brain proteins following injury. *Neurochem. Res.* **22**:1453-1459.
66. **Wilkinson, K., and M. Hochstrasser.** 1998. The deubiquitinating enzymes. pp. 99-126. *In* J.-M. Peters, J. R. Harris, and D. Finley (eds.), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY.

## Legends to Figures

**FIGURE 1.** Deletion/disruption of the mouse *Ntan1* gene. (A) Comparison of enzymatic reactions that underlie the activity of tertiary and secondary destabilizing residues in the yeast *S. cerevisiae* and the mouse. N-terminal residues are indicated by single-letter abbreviations for amino acids. The ovals denote the rest of a protein substrate. The *Ntan1*-encoded mammalian Nt<sup>N</sup>-amidase converts N-terminal Asn into Asp. N-terminal Gln is deamidated by Nt<sup>Q</sup>-amidase, which remains to be isolated (see the main text). In contrast, the yeast Nt-amidase Nta1p can deamidate either N-terminal Asn or Gln (6). The secondary destabilizing residues Asp and Glu are arginylated by the mammalian ATE1-1p or ATE1-2p R-transferases (34). A Cys-specific mammalian R-transferase (23) remains to be identified. N-terminal Arg, one of the primary destabilizing residues, is recognized by N-recognition, the E3 component of the N-end rule pathway (63). (B) The targeting strategy. Top, a partial restriction map of the mouse *Ntan1* gene; middle, the structure of the targeting vector; bottom, the structure of the deletion/disruption *Ntan1*<sup>-/-</sup> allele. Exons are denoted by solid vertical bars. The directions of transcription of the neomycin (*neo*) and the thymidine kinase (*tk*) genes are indicated. Homologous recombination resulted in the replacement of the *Ntan1* exons 2-5 with the *neo* cassette. Probes for Southern hybridization are indicated by solid rectangles. Restriction sites: Xh, *Xho*I; R, *Eco*RI; BI, *Bam*HI; H, *Hind*III; P, *Pst*I. (C) Southern analysis of *Bam*HI-digested tail DNA from wildtype (+/+) and heterozygous (*Ntan1*<sup>+/−</sup>) mice. The 5'-probe yielded the 12 and 1.7 kb *Ntan1* fragments for the wildtype and mutant *Ntan1* alleles, respectively; the 3'-probe detected 12 and 7 kb fragments. The organization of the deletion/disruption allele was independently verified by Southern analysis of the *Xho*I-*Hind*III-digested tail DNA (data not shown). (D) PCR analysis of tail DNA. The primers were 5'-GCCACTTGTGTAGCGCCAAGTGCCAGC (for *neo*, forward), 5'-CTTCCCCACCAAGCCTGACTGTTGATC (for *Ntan1*, forward) and 5'-CTTCAAATTCTGTGCTCAGCTAAGCTC (for *Ntan1*, reverse). (E) RT-PCR analysis of the total RNA isolated from +/+ and *Ntan1*<sup>−/−</sup> EF cells, using primers P1 (for exon 1), P2 (exon 2), P3 (exon 6), P4 (exon 5), and P5 (exon 10). β-actin mRNA was

used as a control, at the 20-fold lower primer concentration in comparison to other lanes.

FIGURE 2. Northern hybridization analysis of the total RNA isolated from +/+ and *Ntan1*<sup>-/-</sup> brains and testes. (Aa-Af) Hybridization using probes a-f that encompassed different regions of the *Ntan1* cDNA (indicated in C). (Ba-Bc) The same Northern blots were hybridized with probes specific, respectively, for the mouse *Ubr1*, *Ate1* and β-actin cDNAs. (C) Exons of *Ntan1* and the hybridization probes used. The sequence of a 208 bp segment of the *Ntan1* cDNA (nt 896-930), termed the IL2 homology region, is 98.6% identical to the sequence of a 206 bp segment in the 3'-flanking untranslated region of the mouse *Il2* gene that encodes interleukin-2 (IL-2) (24).

FIGURE 3. Expression of *Ntan1* mRNA and localization of NTAN1p.

(A) Whole-mount *in situ* hybridization of wildtype and *Ntan1*<sup>-/-</sup> embryos (left 4 panels, light background) with an antisense RNA probe derived from a 0.3 kb fragment of the *Ntan1* cDNA (nt 108-448) that was absent from the *Ntan1*<sup>-/-</sup> allele. The regions of high *Ntan1* expression in the tail buds (t), forelimb buds (fl), and hindlimb buds (hl) are indicated. The right two panels (dark background) show the results of *in situ* hybridization with an antisense (AS) *Ubr1* cDNA probe.

(B) Intracellular localization of mouse NTAN1p. NIH-3T3 cells were transiently transfected with a plasmid expressing the NTAN1p-GFP fusion protein (see Materials and methods). Typical fluorescence patterns and the matching phase contrast images are shown, respectively, in subpanels a, c, e and b, d, f (see the main text).

FIGURE 4. Mouse *Ntan1*<sup>-/-</sup> cells lack Nt<sup>N</sup>-amidase and are unable to degrade the normally short-lived N-end rule substrates bearing N-terminal Asn. (A) <sup>35</sup>S-labeled, purified, X-DHFR test proteins (X=Asn, Gln or Asp) were incubated for 2 hr at 37°C with buffer alone (negative controls, NC) or with extracts from either wildtype or *Ntan1*<sup>-/-</sup> embryonic fibroblasts (EFs), followed by isoelectric focusing (IEF) and autoradiography. The assays were carried using either the initial extracts (lanes a) or the same extracts diluted with buffer by 10-, 100-, and 1,000-fold (lanes, b, c, and d,

respectively). The IEF positions of X-DHFRs bearing N-terminal Asp or Glu versus Asn or Gln are shown on the left. The corresponding pH values are indicated on the right. (B) Immortalized +/+ and *Ntan1*<sup>-/-</sup> EF cells (see Materials and methods) were transiently transfected with the plasmids pRC/dhaUbXnsP4 $\beta$ gal, which expressed DHFR-ha-Ub<sup>R48</sup>-X-nsP4 $\beta$ gal test proteins (X=Met, Asn, Gln, or Arg) (40). These proteins were cotranslationally cleaved *in vivo* by deubiquitylating enzymes (DUBs), yielding the long-lived reference protein DHFR-ha-Ub<sup>R48</sup> and a test protein X-nsP4 $\beta$ gal (X=Met, Asn, Gln, or Arg) (see the main text). Cells were labeled with <sup>35</sup>S-methionine/cysteine, followed by a chase for 0, 1, and 2 hr (as indicated at the top of the panel) in the presence of cycloheximide, preparation of extracts, immunoprecipitation, SDS-PAGE, autoradiography and quantitation, essentially as described (40). The bands of X-nsP4 $\beta$ gal proteins and the DHFR-ha-Ub<sup>R48</sup> reference protein are indicated on the right as X- $\beta$ gal and DHFR. (C) Same as B but with immortalized *Ntan1*<sup>-/-</sup> EF cells. (D) Quantitation of the *in vivo* degradation of X-nsP4 $\beta$ gal test proteins using the reference-based pulse-chase patterns in B and C (see Materials and methods). The amounts of <sup>35</sup>S in an X-nsP4 $\beta$ gal protein, relative to <sup>35</sup>S in the DHFR-ha-Ub<sup>R48</sup> reference protein at the same time points, were plotted as percentages of this ratio for Met-nsP4 $\beta$ gal at time 0. (Met-nsP4 $\beta$ gal bore a stabilizing N-terminal residue.) Open and closed symbols: wildtype and *Ntan1*<sup>-/-</sup> EF cells, respectively. □, ■: Met-nsP4 $\beta$ gal; ○, ●: Asn-nsP4 $\beta$ gal; ▽, ▼: Arg-nsP4 $\beta$ gal; △, ▲: Gln-nsP4 $\beta$ gal.

FIGURE 5. Normal motor coordination and reduced spontaneous activity of *Ntan1*<sup>-/-</sup> mice. Open symbols: +/+ mice. Solid symbols: congenic (same-strain background) *Ntan1*<sup>-/-</sup> mice (see Materials and methods). Statistically significant differences ( $p<0.05$ ) are indicated by \*. 24 *Ntan1*<sup>-/-</sup> mice and their congenic +/+ littermates were used in the experiments of A-C. Non-littermates were used in the experiments of D-G. (A) The rotarod test. Time elapsed before the animals fell from a horizontal rod rotating at 10 rpm (squares) or 20 rpm (circles). No significant differences between +/+ and *Ntan1*<sup>-/-</sup> mice. (B) Weight retention test. Time elapsed

before the animals released a hook. The height of each bar represents the time averaged from 360 trials. No significant differences between +/+ and *Ntan1*<sup>-/-</sup> mice. (C) Coat hanger test. Squares: time elapsed before the animal grabbed the horizontal bar with both rear paws. Circles: time elapsed before the front two paws reached one of the side bars. Each data point is an average of 48 trials. No significant differences between +/+ and *Ntan1*<sup>-/-</sup> mice. (D) Shuttlebox avoidance test. Number of null responses per 50 daily trials in which the mouse remained in the original compartment of shuttle box and received 20 sec of electric shock during the original learning and retention testing 7-10 weeks later. (E) Passive avoidance test. Time before entering dark chamber after an automated guillotine door opened, exposing the dark chamber. On Day 1 the mice were shocked after entering the dark chamber. 13 *Ntan1*<sup>-/-</sup> mice and 12 congenic +/+ mice were used. (Fa-Fd) Open field test. (Fa) The total distance traveled by the animal during the observation time of 9 min. (Fb) Time (out of 9 min total) that the animal spent in the center area. (Fc) Time (out of 9 min total) that the animal spent within 1 cm of the walls. 28 *Ntan1*<sup>-/-</sup> mice and 27 congenic +/+ mice were used in this set of tests.

FIGURE 6. Socially conditioned differences in the exploratory behavior between +/+ and congenic *Ntan1*<sup>-/-</sup> mice. Open bars, +/+ mice; solid bars, congenic (same-strain background) *Ntan1*<sup>-/-</sup> mice. Platform-leaving test: one *Ntan1*<sup>-/-</sup> mouse and one +/+ mouse (either a littermate or a non-littermate) were placed, at the same time, on the 16 x 22 x 2.7 cm platform. Mice were allowed to either explore or step down from the platform until the cutoff time of 180 sec. (Aa-Ac) Pairs of littermates. The time by which the animal left the platform (a); the number of animals, of each genotype, leaving first (b); and the number of animals, of each genotype, not leaving by the cutoff time (c). Four independent sets of tests (at least 10 pairs of mice per test) were carried out over 6 months, using 44 pairs of +/+ and *Ntan1*<sup>-/-</sup> littermates, which were produced through heterozygous (+/-) matings and identified by genotyping ~700 mice. The results differed by less than 15% among the independent tests. (Ba-Bc) Same as A except with pairs of non-littermates. 24 +/+ mice and 24 *Ntan1*<sup>-/-</sup>

mice were examined in this test.

FIGURE 7. Spatial learning and memory in *Ntan1*<sup>-/-</sup> mice. Open symbols or bars represent +/+ mice. Solid symbols or bars represent congenic (same-strain background) *Ntan1*<sup>-/-</sup> mice. See the main text and Materials and methods for detail. 10 +/+ mice and 10 *Ntan1*<sup>-/-</sup> mice were examined in the tests of A-C. (A) Mean daily errors made by mice for each phase in the spatial radial arm maze. The data were separated into those for acquisition phase (sessions 2-7) and asymptotic phase (sessions 8-12). (B) Same as A except that the mice were tested in the non-spatial radial arm maze, in which the spatial cues were replaced by non-spatial cues. (C) Effects of memory load in the water version of the spatial radial-arm maze. Shown are total errors per trial. (Da-Dc) Performance of +/+ (open circles; total of 29 mice) and congenic *Ntan1*<sup>-/-</sup> mice (solid circles; total of 30 mice) in the Lashley maze. The ratio of the number of correct path segments taken to the total number of segments (learning index) (a); errors in the T-choices (making the correct left or right turn) (b); errors made by moving away from the goal (c).

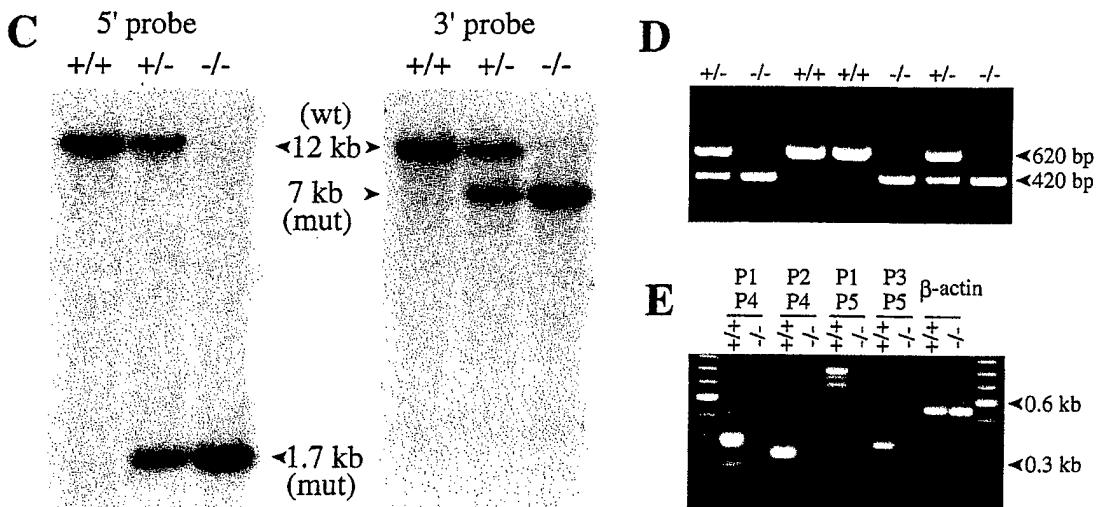
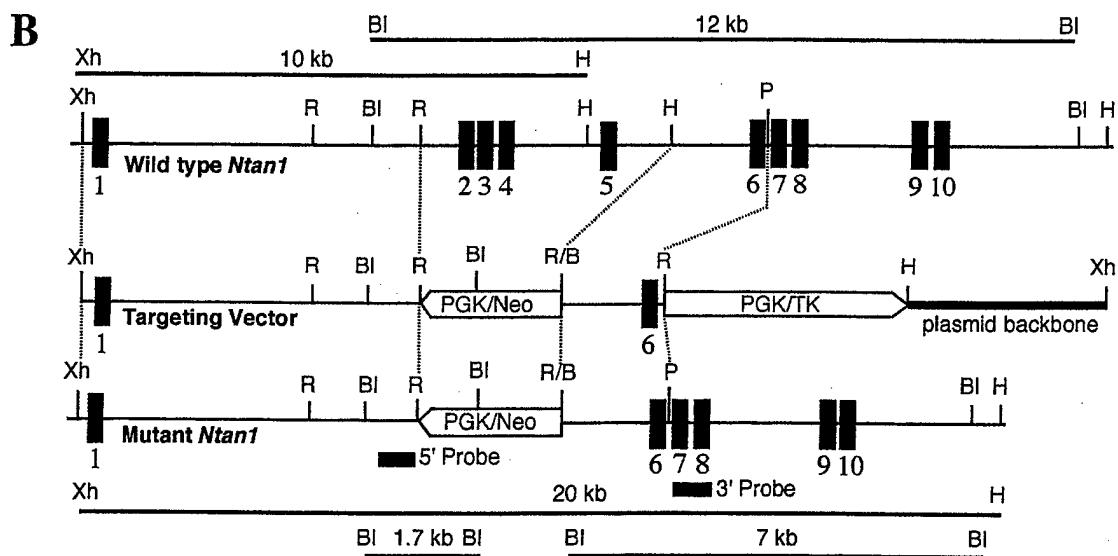
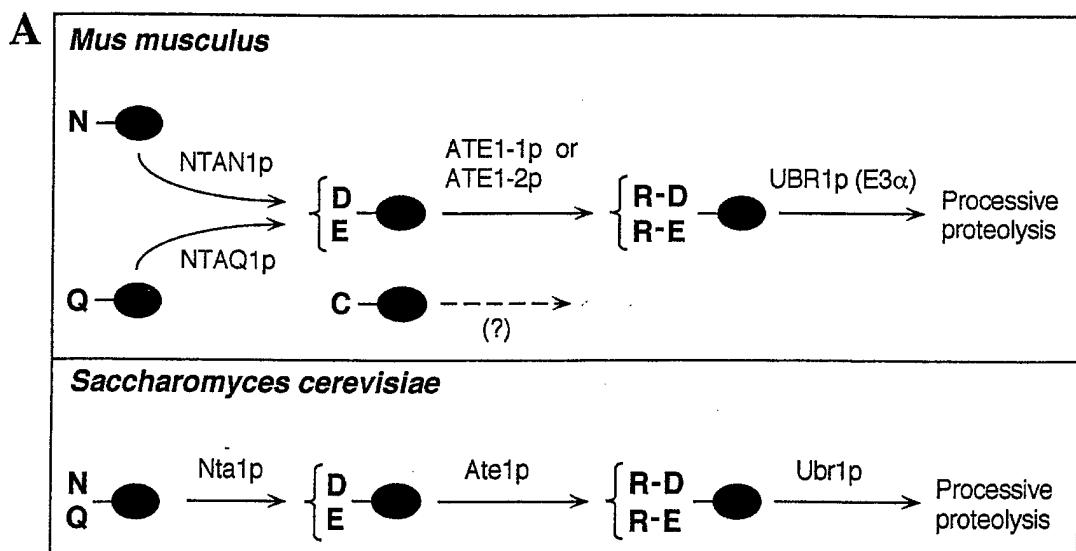


FIGURE 1

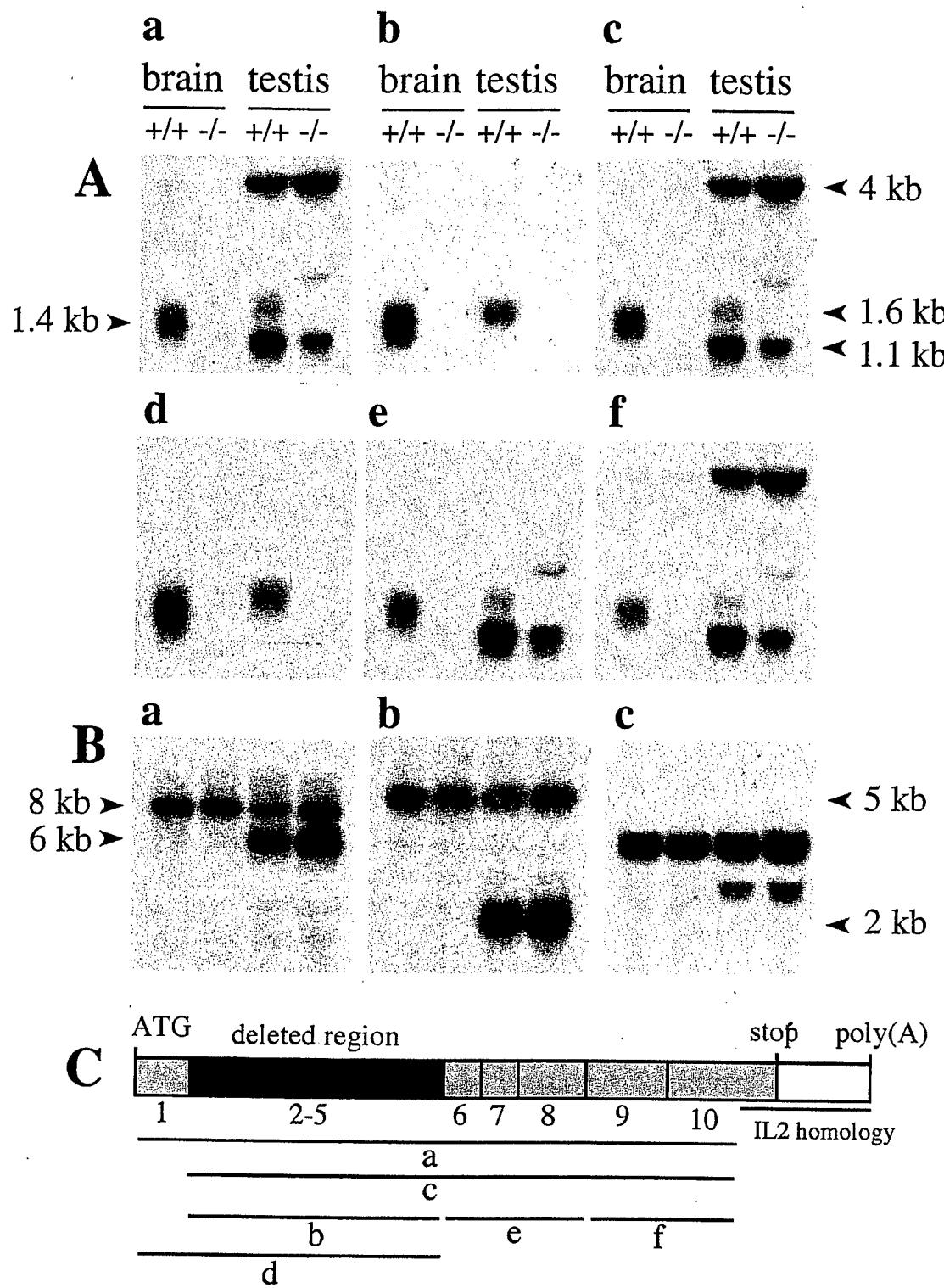
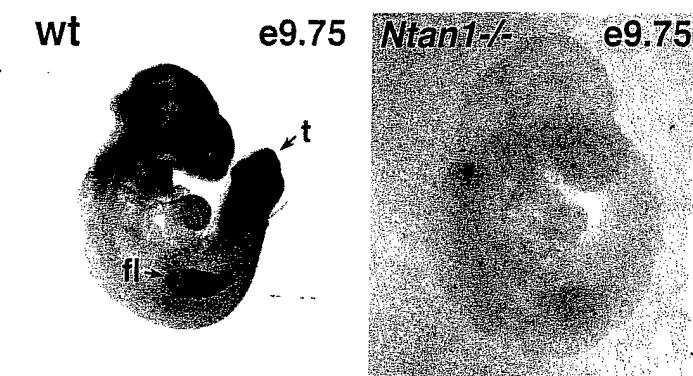


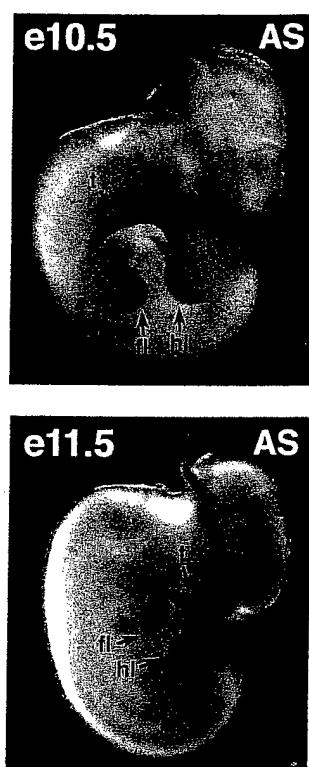
FIGURE 2

## *Ntan1*

**A**



## *Ubr1*



**B**

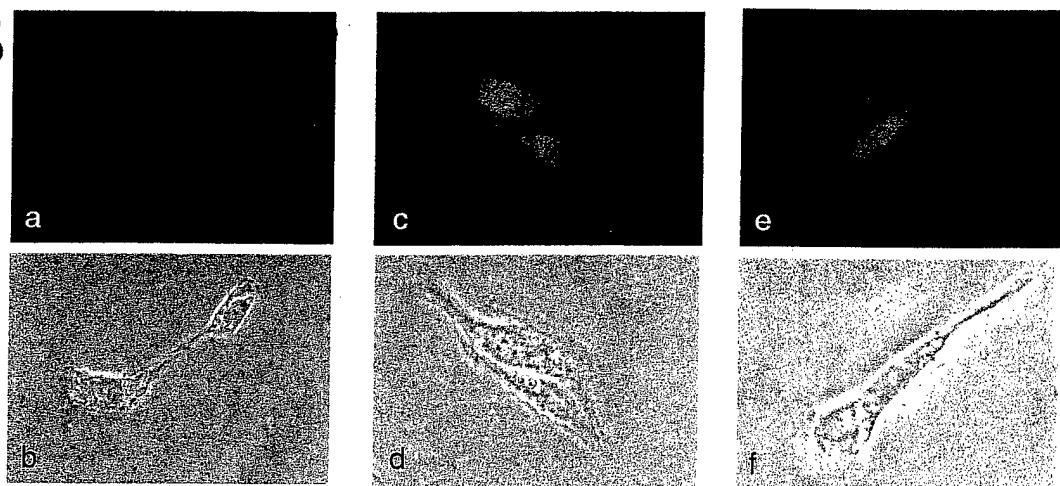


FIGURE 3

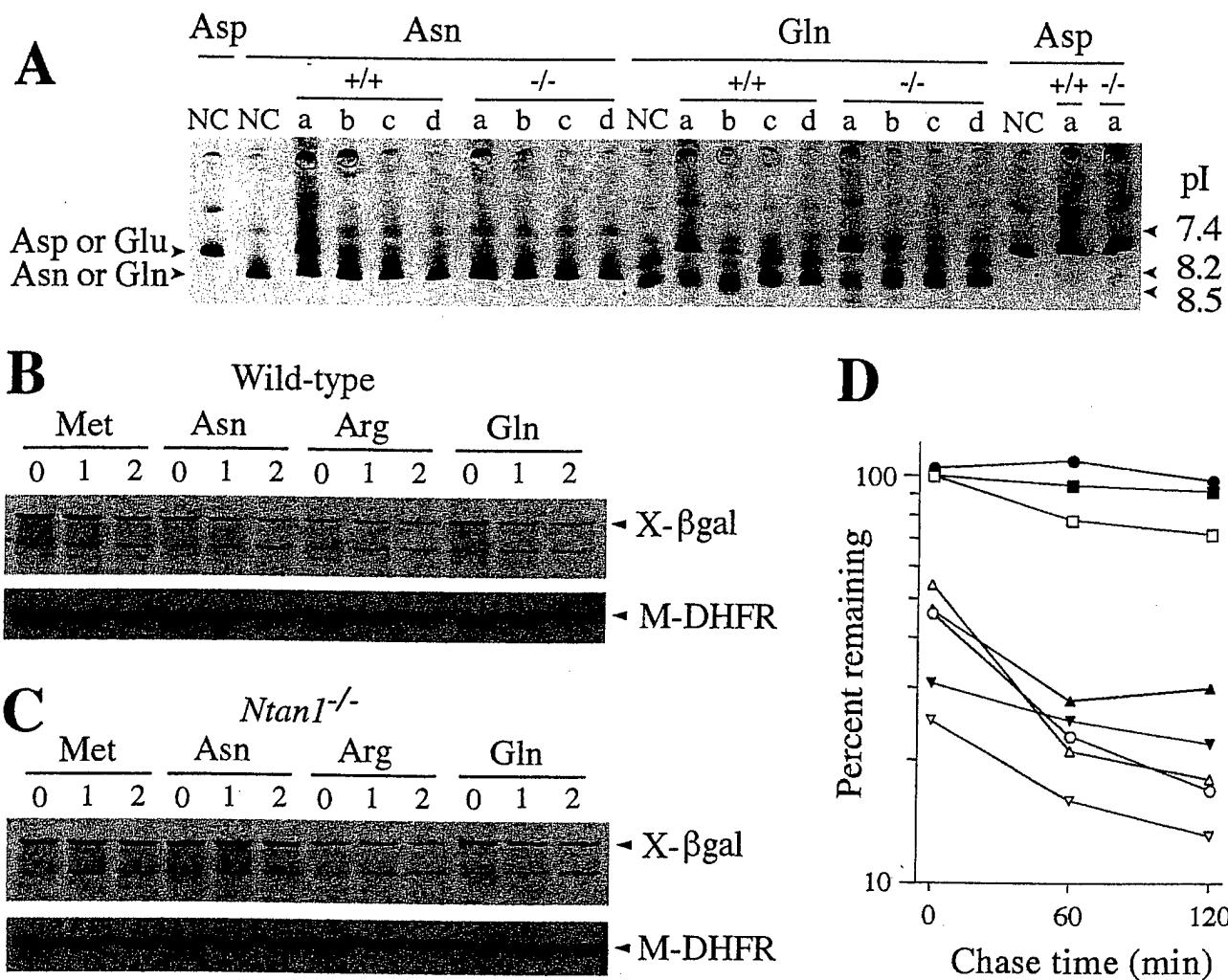


FIGURE 4

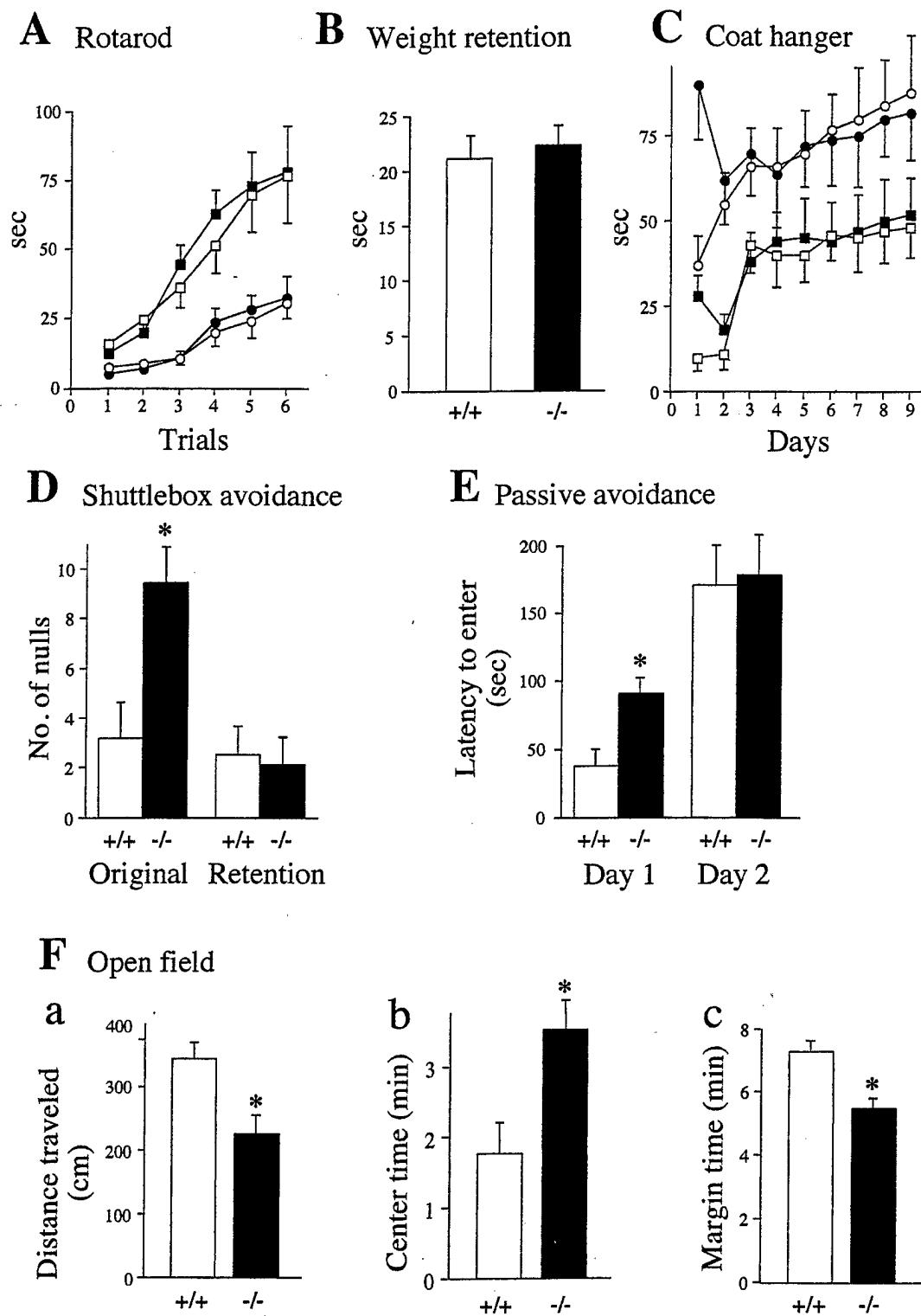
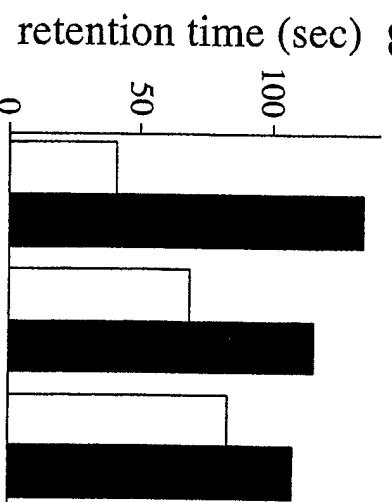


FIGURE 5

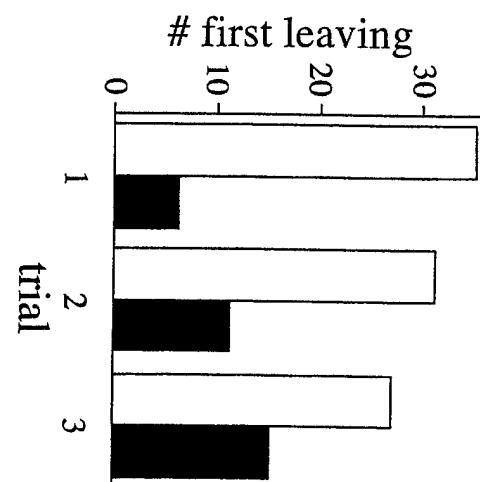
**A Littermates**



**B Non-littermates**



**b**



**c**

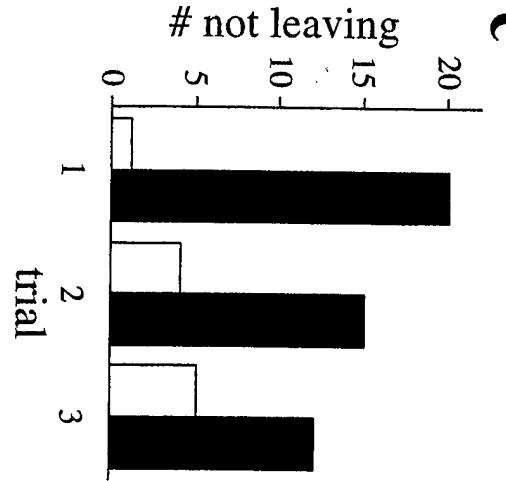


FIGURE 6

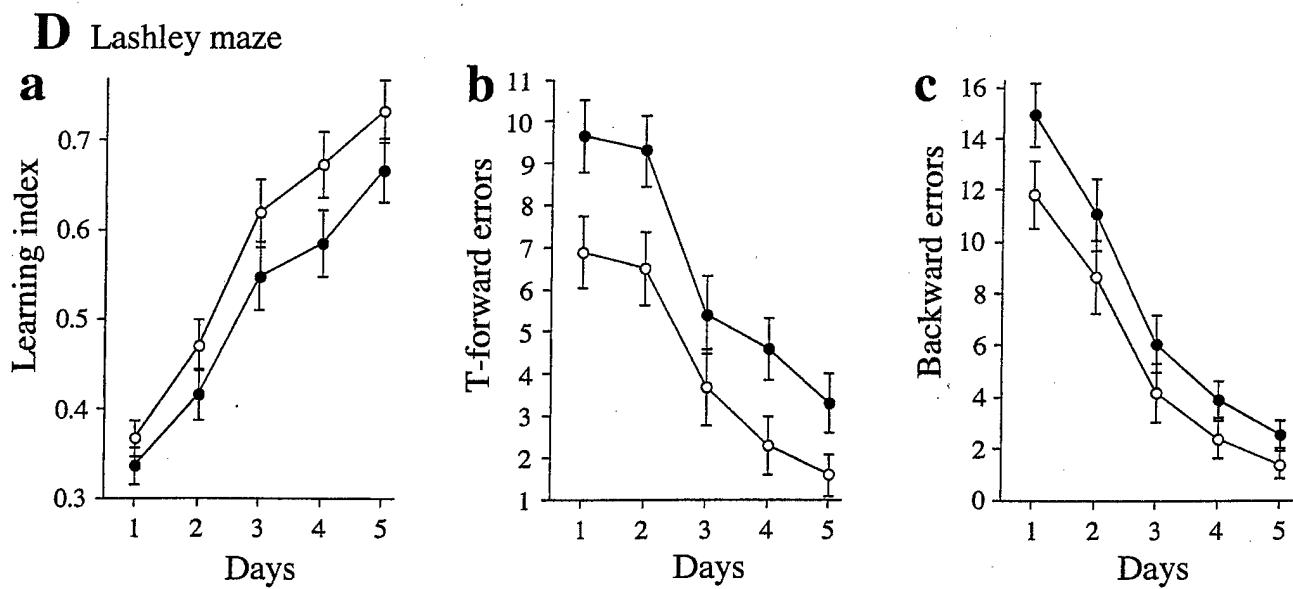
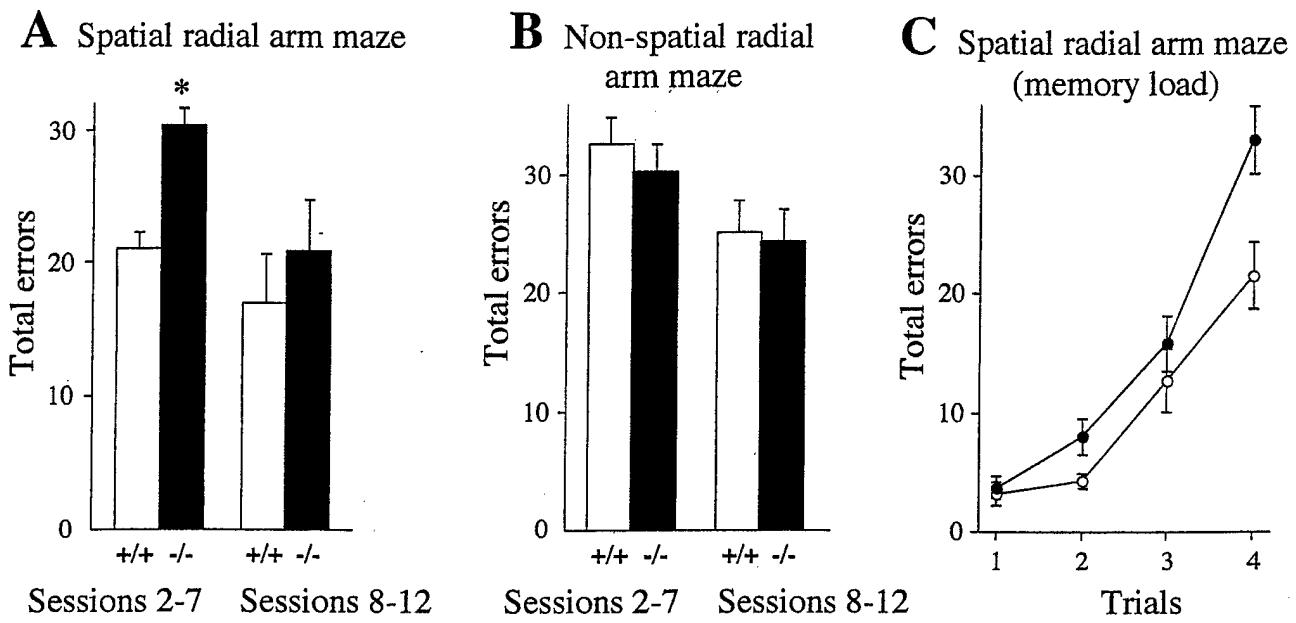


FIGURE 7

## Detecting and Measuring Cotranslational Protein Degradation *In Vivo*

Glenn C. Turner and Alexander Varshavsky\*

Division of Biology,  
California Institute of Technology,  
Pasadena, CA 91125, USA

\*Corresponding author: Alexander Varshavsky, Division of Biology, 147-75,  
Caltech, 1200 East California Boulevard, Pasadena, CA 91125, USA

Telephone: 626-395-3785; Fax: 626-440-9821; Email: avarsh@cco.caltech.edu

*Keywords:* proteolysis / ubiquitin / proteasome / N-end rule / pulse-chase /

**Abstract**

Nascent polypeptides emerging from the ribosome and not yet folded may at least transiently present degradation signals similar to those recognized by the ubiquitin system in misfolded proteins. Thus, the folding of nascent proteins, including abnormal ones, may be in kinetic competition with pathways that target these proteins for degradation cotranslationally. Here we describe the ubiquitin sandwich technique that allows the detection and measurement of cotranslational protein degradation in living cells. We use this method to demonstrate that more than 50% of nascent protein molecules bearing an N-terminal degradation signal can be degraded cotranslationally, never reaching their mature size before their destruction by processive proteolysis.

Nascent (being synthesized) polypeptides emerging from the ribosome may, in the process of folding, present degradation signals similar to those recognized by the ubiquitin system in misfolded or otherwise damaged proteins (1). It has been a long-standing question whether a significant fraction of nascent polypeptides is cotranslationally degraded (2). Determining whether nascent polypeptides are actually degraded *in vivo* has been difficult because at any given time the nascent chains of a particular protein species are of different sizes, and therefore would not form a band upon electrophoresis in a conventional pulse-chase assay (3). The ubiquitin (Ub) sandwich technique makes it possible to detect cotranslational protein degradation by measuring the steady-state ratio of two reporter proteins whose relative abundance is established cotranslationally.

The polypeptide to be examined for cotranslational degradation, termed **B**, is sandwiched between two stable reporter domains **A** and **C** in a linear fusion protein. The three polypeptides are connected via Ub moieties to create a fusion protein of the form **A**Ub-**B**Ub-**C**Ub (Fig. 1A). Ub-specific processing proteases (UBPs) (4) cotranslationally cleave such linear Ub fusions after the last residue of Ub (5-8). The independent polypeptides **A**Ub, **B**Ub and **C**Ub that result from the cleavage of **A**Ub-**B**Ub-**C**Ub are called modules below.

UBP-mediated cleavage establishes a kinetic competition between two mutually exclusive events during the synthesis of the **A**Ub-**B**Ub-**C**Ub fusion: cotranslational UBP cleavage at the **B**Ub-**C**Ub junction to release the long-lived **C**Ub module or, alternatively, cotranslational degradation of the entire **B**Ub-**C**Ub nascent chain by the 26S proteasome (9) (Fig. 1B). In the latter case, the processivity of proteasome-mediated degradation results in the destruction of the Ub moiety between **B** and **C** before it can be recognized by UBPs. The resulting drop in levels of the **C**Ub module relative to levels of **A**Ub, referred to as the C/A ratio, reflects the cotranslational degradation of domain **B** (Fig. 1B). This measurement provides a *minimal* estimate of the total amount of cotranslational degradation, because non-processive cotranslational degradation events that do not extend into the **C** domain are not detected.

It should be noted that the Ub moieties of the fusion serve solely as cotranslationally cleavable junctions between domains **A**, **B**, and **C**. They do not target the attached proteins for degradation (8). In addition, a lysine to arginine substitution at position 48 of these moieties ( $\text{Ub}^{\text{R48}}$ ) prevents their conjugation to other Ub molecules at that position, so they cannot be involved in the formation of a major class of substrate-linked multi-Ub chains required for degradation by the 26S proteasome (10, 11).

UBP-mediated cleavage was previously used to examine the *in vivo* kinetics of protein translocation into the endoplasmic reticulum (ER) (6), where it was shown that cleavage was fast enough to be cotranslational. We directly verified this in the case of an **A**Ub-**B**Ub-**C**Ub fusion, using an *in vivo* radiolabeling regimen in which the labeling pulse was significantly shorter than the time required for the complete synthesis of **A**Ub-**B**Ub-**C**Ub, and terminating the pulse by arresting translation. At the beginning of the pulse, the cell's ribosomes are at various stages in the synthesis of the fusion. As a result, the pulse produces a population of nascent **A**Ub-**B**Ub-**C**Ub chains that are labeled at various regions along the chain (Fig. 2B). The nascent chains that are just starting to be synthesized when the pulse begins will incorporate label into the N-terminal **A** domain, but they will not be elongated to full-length chains before the end of the pulse. Under these conditions free *labeled* **A**Ub could be produced only if UBP cleave the nascent polypeptide chains (Fig. 2B).

*S. cerevisiae* cells expressing a 190 kD **A**Ub-**B**Ub-**C**Ub of the form {DHFRhaUb} - {Me<sup>K</sup>βgalUb} - {Me<sup>K</sup>DHFRhaUb} (Fig. 2A) were labeled with <sup>35</sup>S-methionine/cysteine for 45 sec (12). In the eukaryotic cell types examined, rates of translation vary between 2-10 residues/sec, while a rate of 12 residues/sec has been measured in *E. coli* (13). Assuming a moderate estimate of 5 residues/sec in *S. cerevisiae*, the above fusion would require ~350 sec for complete synthesis, more than 7 times the duration of the 45 sec pulse. The labeling was terminated by the addition of the translation inhibitor cycloheximide, and UBP were simultaneously inactivated with N-ethylmaleimide (NEM) (6). Discrete bands corresponding to the labeled **A**Ub (DHFRhaUb) and **B**Ub (Me<sup>K</sup>βgalUb) modules

were observed upon immunoprecipitation, indicating that UBP-mediated cleavage at the AUb-BUb junction was cotranslational (Fig. 2C). Moreover, no full-length AUb-BUb-CUb fusion was detected (Fig. 2C), indicating that the cotranslational cleavage by UBPs was highly efficient. High concentrations of NEM in these extracts increased nonspecific immunoprecipitation, yielding a crossreacting species that co-migrated with CUb ( $\text{Me}^K\text{DHFRhaUb}$ ). No such crossreacting band was present when NEM was omitted (Fig. 3A). However, only the presence of free labeled AUb was diagnostic of cotranslational cleavage; free CUb was expected to be observed whether cleavage by UBPs was post- or cotranslational (Fig. 2B).

This result confirmed the key assumption of the Ub sandwich technique, validating its use to detect cotranslational degradation *in vivo*. Previous evidence bearing on cotranslational degradation was based on experiments with inhibitors or cell-free systems (14). Nascent polypeptide chains might be protected from degradation *in vivo*, either because they are sterically shielded by chaperones or because their translation time is short in comparison to the time required for targeting by the degradation machinery. To maximize the likelihood of detecting cotranslational degradation, a large, 118 kD,  $\beta$ gal-derived polypeptide carrying a strong N-terminal degradation signal was initially chosen as the B domain of the AUb-BUb-CUb fusion (Fig. 2A). A protein of this size is unable to fit into the cavity of the cytosolic CCT chaperonin (15), ensuring that this type of shielding was unavailable to the test protein. In addition, the N-terminal location of the degradation signal made it potentially accessible from the beginning of translation.

The degradation signal used was an N-degron, which is targeted by Ubr1p, the E3 component of the N-end rule pathway (5, 16). The  $\beta$ gal-linked N-degron comprises the destabilizing N-terminal residue arginine (R) and a short, lysine-bearing extension, e<sup>K</sup> (Fig. 2A) that is the site of multi-Ub chain attachment (11, 17). Re<sup>K</sup>- $\beta$ gal is rapidly degraded *in vivo* ( $t_{1/2} \sim 2$  min) (5, 7). Changing just the N-terminal residue of the protein to methionine (M) inactivates the degradation signal by precluding recognition by Ubr1p. The resulting Me<sup>K</sup>- $\beta$ gal is posttranslationally stable ( $t_{1/2} > 20$  hr) (5, 7). AUb-BUb-CUb fusions in which domain B was either Re<sup>K</sup>- $\beta$ gal or its N-degron-lacking counterpart Me<sup>K</sup>- $\beta$ gal

(Fig. 2A) were expressed in *S. cerevisiae* strains containing different levels of Ubr1p (18). The extent of cotranslational degradation was assessed by radiolabeling for 30 min and immunoprecipitation (19) to determine the levels of CUb relative to AUb; this parameter is referred to below as the C/A ratio.

The results of a representative experiment are shown in Fig. 3A. The C/A ratio was significantly and reproducibly lower in cells expressing the N-degron-bearing domain **B**, but only in those strains that also expressed Ubr1p. To determine the percentage of nascent chains cotranslationally degraded by the N-end rule pathway, we averaged the C/A ratios over several trials, and compared the ratios for Re<sup>K</sup>- $\beta$ gal in wildtype (0.52) and Ubr1p-overexpressing strains (0.39) to the ratio found with the *ubr1* $\Delta$  strain (0.86) (Fig. 4A). This comparison indicated that ~40% of the nascent Re<sup>K</sup>- $\beta$ gal chains were cotranslationally degraded in the wildtype strain (Fig. 4B) (20). This fraction increased to ~55% when Ubr1p was overexpressed from the P<sub>GAL1</sub> promoter.

A different way to measure the extent of N-degron-dependent cotranslational degradation was to compare the C/A ratios for Re<sup>K</sup>- $\beta$ gal and Me<sup>K</sup>- $\beta$ gal in the same strain. This comparison *independently* indicated ~40% cotranslational degradation of Re<sup>K</sup>- $\beta$ gal in the wildtype strain (R = 0.52; M = 0.82), and ~55% when Ubr1p was overexpressed (R = 0.39; M = 0.87). Note that the C/A ratios for Re<sup>K</sup>- $\beta$ gal in the *ubr1* $\Delta$  strain, and for Me<sup>K</sup>- $\beta$ gal in all the strains used, were ~0.85 (Fig. 4A), significantly less than the 1.0 ratio expected in the absence of cotranslational degradation; this is discussed below.

The extent of cotranslational degradation was expected to depend on the length of the **B** domain (Fig. 1B), because a larger protein requires a longer time for synthesis, and possibly also because the nascent chain of a larger protein extends a greater distance from the ribosome. To address this issue, we determined the extent of cotranslational degradation of two other **B** domains. The mammalian Sindbis virus RNA polymerase, termed nsP4, is a 69 kD protein that naturally bears an N-degron (21). Xe<sup>K</sup>-Ura3p (X=M or R) is a 34 kD enzyme of the *S. cerevisiae* uracil biosynthetic pathway that either carried (Re<sup>K</sup>-Ura3p) or lacked (Me<sup>K</sup>-Ura3p)

an N-degron (22) (Fig. 2A). Initially, immunoblot analysis was used to estimate the steady-state C/A ratios for these AU<sub>B</sub>-BU<sub>B</sub>-CU<sub>B</sub> fusions (Fig. 3B) (23). While the short-lived R-nsP4 (69 kD) and Re<sup>K</sup>-βgal (118 kD) were cotranslationally degraded to similar extents, Re<sup>K</sup>-Ura3p (34 kD), which was also short-lived posttranslationally (22) exhibited much less cotranslational degradation than the other two proteins (Fig. 3B). To measure C/A ratios more accurately, *in vivo* radiolabeling and immunoprecipitations were carried out. The results of several experiments, summarized in Figure 4, clearly showed that in the presence of a strong N-degron, cotranslational degradation is significant for all three of the tested proteins, Re<sup>K</sup>-βgal, R-nsP4, and Re<sup>K</sup>-Ura3p.

Comparisons of the C/A ratios for R-nsP4 in the wildtype (*UBR1*) and *ubr1Δ* strains indicated that ~50% of the nascent chains were cotranslationally degraded by the N-end rule pathway in wildtype cells, and ~55% in cells overexpressing Ubr1p (Fig. 4B). These values were similar to, respectively, ~40% and ~55% cotranslational degradation of Re<sup>K</sup>-βgal in these strains (Fig. 4B). In contrast, only ~20% cotranslational degradation was observed with the 34 kD Re<sup>K</sup>-Ura3p in either wildtype or Ubr1-overexpressing strains (Fig. 4B), suggesting that smaller proteins are less susceptible to cotranslational degradation.

These comparisons established the levels of cotranslational degradation by the Ubr1p-dependent N-end rule pathway. In addition, the C/A ratios for two of the N-degron-lacking B domains, Me<sup>K</sup>-βgal and M-nsP4, were less than 1.0, suggesting that they might be cotranslationally degraded by a Ubr1p-independent pathway. To determine the baseline C/A ratio for negligible cotranslational degradation, we constructed an AU<sub>B</sub>-BU<sub>B</sub>-CU<sub>B</sub> fusion in which domain B was an 18-residue sequence containing the FLAG epitope (Fig. 2A). This sequence bore no known degradation signals and would presumably be too short to be degraded cotranslationally. The C/A ratio observed was 0.97. Thus Me<sup>K</sup>-βgal exhibited a ~15% drop in C/A relative to the FLAG-bearing peptide, M-nsP4 a ~25% drop, whereas the C/A ratios obtained with Me<sup>K</sup>-Ura3 and FLAG were indistinguishable (Fig. 4A). One interpretation of these differences is that, although Me<sup>K</sup>-βgal and

M-nsP4 are posttranslationally stable, they may be cotranslationally degraded to a significant extent. Premature termination of translation and/or transcription may also contribute to the drop in the C/A ratios for these two N-degron-lacking **B** domains. It is unlikely that premature termination of translation entirely accounts for the observed difference between the C/A ratios for FLAG and M-nsP4, because the codon adaptation index of the M-nsp4 open reading frame (0.1) is higher than the one for Me<sup>K</sup>-βgal (0.07), but the drop in C/A ratio is greater with M-nsP4 (25%) than with Me<sup>K</sup>-βgal (15%).

This work demonstrates that a nascent polypeptide carrying a degradation signal can be cotranslationally degraded *in vivo* (Fig. 4B). Cotranslational protein degradation, which cannot be detected by conventional pulse-chase assays, can now be studied using the Ub sandwich technique (Fig. 1). This method revealed that the extent of cotranslational degradation can be strikingly high: in the case of Re<sup>K</sup>-βgal in Ubr1p-overexpressing cells, over 50% of nascent polypeptide chains never reach their full size before their destruction by processive proteolysis. Thus, if a nascent chain displays a degron of the Ub system, such a protein becomes a target of kinetic competition between cotranslational biogenesis and cotranslational degradation. Since the folding of a protein molecule begins during its synthesis on the ribosome, a nascent polypeptide may cotranslationally expose degradation signals in the form of, for example, hydrophobic patches which become shielded through the folding of the newly formed protein (1, 24). Our findings with the posttranslationally long-lived Me<sup>K</sup>-βgal and M-nsP4 (Fig. 4A) are consistent with this possibility.

If cotranslational protein degradation by the Ub system is found to be extensive *in vivo*, it could be accounted for as an evolutionary trade-off between the necessity of identifying and destroying degron-bearing mature proteins and the mechanistic difficulty of distinguishing between posttranslationally and cotranslationally presented degrons. Conditional degradation signals that require posttranslational phosphorylation for activation, such as those in IκBα and β-catenin (25), may have the additional feature of not contributing to cotranslational proteolysis. Cotranslational protein degradation may also represent

a previously unrecognized form of protein quality control, which destroys nascent chains that fail to fold correctly. These and other questions about physiological aspects of cotranslational protein degradation can now be addressed directly in living cells through the Ub sandwich technique.

## References and Notes

1. M. Hochstrasser, *Annu. Rev. Genet.* **30**, 405-439 (1996); A. Varshavsky, *Trends Biochem. Sci.* **22**, 383-387 (1997); A. Hershko and A. Ciechanover, *Annu. Rev. Biochem.* **76**, 425-479 (1998).
2. B. Alberts, et al., *Molecular Biology of the Cell*, p. 220. (Garland Publishing, New York, 1994).
3. J.-M. Peters, J. R. Harris, D. Finley, Eds., *Ubiquitin and the Biology of the Cell*. (Plenum Press, New York, NY, 1998).
4. K. Wilkinson and M. Hochstrasser, in *Ubiquitin and the Biology of the Cell*. J.-M. Peters, J. R. Harris, D. Finley, Eds. (Plenum Press, New York, NY, 1998).
5. A. Bachmair, D. Finley, A. Varshavsky, *Science* **234**, 179-186 (1986).
6. N. Johnsson and A. Varshavsky, *EMBO J.* **13**, 2686-2698 (1994).
7. A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **93**, 12142-12149 (1996).
8. F. Lévy, N. Johnsson, T. Rumenapf, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **93**, 4907-4912 (1996).
9. W. Baumeister, J. Walz, F. Zühl, E. Seemüller, *Cell* **92**, 367-380 (1998); G. N. DeMartino and C. A. Slaughter *J. Biol. Chem.* **274**, 22123-22126 (1999).
10. C. M. Pickart, *FASEB J.* **11**, 1055-1066 (1997); M. Scheffner, S. Smith, S. Jentsch, in *Ubiquitin and the Biology of the Cell* J.-M. Peters, J. R. Harris, D. Finley, Eds. (Plenum Press, New York, 1998) pp. 65-98.
11. V. Chau, et al., *Science* **243**, 1576-1583 (1989).
12. The expression of Ub sandwich proteins from the P<sub>GAL1</sub> promoter was induced with 0.1 μM β-estradiol, using the Gal4.ER.VP16 chimeric protein, which consisted of the DNA-binding domain of Gal4p, the estrogen-binding domain of estrogen receptor, and the VP16 activation domain expressed from a low copy plasmid (26). Induction commenced when cultures were at OD<sub>600</sub> ~0.2. Cells were harvested after ~10 hr of growth to OD<sub>600</sub> 0.5-1. The harvested cells were washed in 0.8 ml SD medium (27) containing 0.1 μM β-estradiol and resuspended in 0.4 ml of the same medium. Cells were labeled for 30 sec with 0.28 mCi of <sup>35</sup>S-EXPRESS (New England

Nuclear). Cycloheximide and NEM (final concentrations 0.5 mg/ml and 0.1 M respectively) were then added, and 0.2 ml of the resulting mixture transferred to a chilled tube containing 0.5 ml of 0.5-mm glass beads, 0.8 ml of ice-cold lysis buffer (1% Triton-X100, 0.15 M NaCl, 5 mM EDTA, 50 mM Na-HEPES, pH 7.5) and a mixture of protease inhibitors (final concentrations 1 mM phenylmethylsulfonyl fluoride, 2 µg/ml aprotinin, 0.5 µg/ml leupeptin, and 0.7 µg/ml pepstatin). Total handling time to the point of beginning the lysis was 45 sec. Extracts were prepared and immunoprecipitations carried out as described (22), using the following monoclonal antibodies, as appropriate: anti-ha 12CA5 (Boehringer), anti-FLAG M2 (Eastman Kodak), and anti-βgal (Promega). Immunoprecipitates were fractionated by SDS-13% PAGE. The bands were detected by autoradiography and quantitated by PhosphorImager (Molecular Dynamics).

13. A. S. Spirin, *Ribosome Structure and Protein Biosynthesis*. (Benjamin Cummings Publ. Co., Menlo Park, CA, USA, 1986); M. A. Sorensen, C. G. Kurland, S. Pedersen, *J. Mol. Biol.* **207**, 365-377 (1989).
14. W. Liao, S. C. J. Yeung, L. Chan, *J. Biol. Chem.* **273**, 27225-27230 (1998); L. Lin, G. N. DeMartino, W. C. Greene, *Cell* **92**, 819-828 (1998); S. Sato, C. L. Ward, R. R. Kopito, *J. Biol. Chem.* **273**, 7189-7192 (1998).
15. A. L. Horwich, K. B. Low, W. A. Fenton, I. N. Hirshfield, K. Furtak, *Cell* **74**, 909-917 (1993); J. Frydman and F. U. Hartl, *Science* **272**, 1497-1502 (1996); F. U. Hartl, *Nature* **381**, 571-580 (1996); J. L. Johnson and E. A. Craig, *Cell* **90**, 201-4 (1997); B. Bukau and A. L. Horwich, *Cell* **92**, 351-366 (1998); R. J. Ellis, *Curr. Biol.* **9**, 137-139 (1999).
16. B. Bartel, I. Wünning, A. Varshavsky, *EMBO J.* **9**, 3179-3189 (1990); K. Madura, R. J. Dohmen, A. Varshavsky, *J. Biol. Chem.* **268**, 12046-12054 (1993).
17. A. Bachmair and A. Varshavsky, *Cell* **56**, 1019-1032 (1989).
18. The *S. cerevisiae* strains JD52 (*MATa lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*), JD54 (*P<sub>GAL1</sub>::UBR1*) and JD55 (*ubr1Δ*) were described previously (22).

19. As in (12), with the following modifications: cells were labeled for 30 min at 30°C with 0.11 mCi of  $^{35}\text{S}$ -EXPRESS. Cells were pelleted, resuspended in 0.8 ml of ice-cold lysis buffer, then transferred to a chilled tube containing protease inhibitors, as described (12).
20. The C/A ratio was the amount of  $^{35}\text{S}$  in the CUb band divided by the amount of  $^{35}\text{S}$  in the AUb band. To determine the level of the Ubr1p-dependent cotranslational degradation, the C/A ratio observed for a particular N-degron-bearing B domain in the wildtype or Ubr1p-overexpressing strain was divided by the C/A ratio for the same B domain in the *ubr1Δ* strain, and the resulting value was subtracted from 100%. For example, with Re<sup>K</sup>-βgal, C/A was 0.52 in wildtype cells, and 0.86 in *ubr1Δ*, so the Ubr1p-dependent cotranslational degradation was  $100\% \times (1 - 0.52/0.86) = 40\%$ .
21. R. J. deGroot, T. Rümenapf, R. J. Kuhn, J. H. Strauss, *Proc. Natl. Acad. Sci. USA* **88**, 8967-8971 (1991).
22. M. Ghislain, R. J. Dohmen, F. Levy, A. Varshavsky, *EMBO J.* **15**, 4884-4899 (1996).
23. Yeast extracts were prepared for immunoblotting as described for immunoprecipitation, except that disruption of unlabeled cells was carried out in 0.2 ml of lysis buffer with 0.2 ml of glass beads. Immunoblotting was carried out with anti-ha at 1:1000 dilution, followed by a secondary horseradish peroxidase-conjugated goat anti-mouse antibody (BioRad) at 1:1000 dilution. The blots were developed using ECL (Amersham).
24. T. Gilon, O. Chomsky, R. G. Kulka, *EMBO J.* **17**, 2759-2766 (1998); W. J. Netzer and F. U. Hartl, *Trends Biochem. Sci.* **23**, 68-73 (1998).
25. Z. J. Chen, et al., *Genes Dev.* **9**, 1586-1597 (1995), J. T. Winston, et al., *Genes Dev.* **13**, 270-283 (1999).
26. J. F. Louvion, B. Havauxcopf, D. Picard, *Gene* **131**, 129-134 (1993).
27. F. M. Ausubel, et al., Eds., *Current Protocols in Molecular Biology*. (Wiley-Interscience, New York, 1996).

28. D. Mumberg, R. Muller, M. Funk, *Nucl. Acids Res.* **22**, 5767-5768 (1994).
29. Plasmids encoding Ub sandwich fusions were constructed using standard PCR and recombinant DNA techniques (27), and the *E. coli* strain DH5 $\alpha$ F'(*recA1*) to minimize recombination events. Construction details are available upon request. All Ub fusion proteins were expressed from the P<sub>GAL1</sub> promoter on the high copy p426GAL1 vector (28).
30. Comparisons of means by the Mann-Whitney test were carried out using InStat 2.00 (Graphpad Software).
31. We thank F. Lévy and N. Johnsson for their advice at early stages of this work. We also thank R. Deshaies, T. Iverson, and members of the Varshavsky laboratory for helpful discussions and comments on the manuscript. This study was supported by a grant to A. V. from the NIH. G. T. was supported in part by Amgen, Inc.

### Figure Legends

FIGURE 1. The ubiquitin sandwich technique. (A) Organization of a Ub sandwich fusion. **B**, the polypeptide assayed for cotranslational degradation, is sandwiched between two stable reporter domains **A** and **C**. Red arrows indicate the locations of UBP cleavage sites. (B) The principle of the method. The reporter module **AUb** is the first to be synthesized, and is cotranslationally released from **B**, thereby providing a measure of the number of nascent **B** chains that initially emerge from the ribosome. If the processive degradation of the emerging **B**, indicated by its insertion into the cylindrical proteasome, is strictly posttranslational, the UBP-mediated cleavage at the **B****Ub**-**C** junction releases **C****Ub** before **B** is degraded, so the molar yields of **C****Ub** and **AUb** are identical. However, if the degradation of **B** can be cotranslational, a significant fraction of **B****Ub**-**C****Ub** may be degraded as a unit. This will result in the molar yield of **C****Ub** being lower than **AUb**, the difference being a measure of cotranslational degradation.

FIGURE 2. UBP-mediated cleavage of ubiquitin sandwich fusions is cotranslational. (A) The protein fusions used (29). Domains **A** and **C** are mouse dihydrofolate reductase tagged with the influenza haemagglutinin-derived ha epitope (DHFRha). Domain **C** carries an N-terminal extension ( $e^K$ , see text) which makes it electrophoretically distinguishable from **A**. The different **B** domains are *E. coli*  $\beta$ -galactosidase ( $\beta$ gal), Sindbis virus RNA polymerase (nsP4) and *S. cerevisiae* Ura3. Unstable (N-degron-bearing) versions of these domains have an N-terminal arginine (R) residue; stable versions have methionine (M) (B) The population of nascent chains produced by a radiolabeling pulse significantly shorter than the time of translation of an **A****Ub**-**B****Ub**-**C****Ub** fusion. Stretches of the polypeptide containing radiolabel are in red; unlabeled stretches are in black. If UBPs efficiently cleave the nascent chain, free radiolabeled **A****Ub**, **B****Ub** and **C****Ub** should all be detected. If UBPs can cleave solely the full-length, mature protein, only the labeled **C****Ub** module will be observed. (C) The UBP cleavage of Ub sandwich fusions is cotranslational. *S. cerevisiae* expressing an **A****Ub**-**B****Ub**-**C****Ub** fusion of the form {DHFRhaUb}- $\{Me^K\beta galUb\}$ - $\{Me^K DHFRhaUb\}$  which is

predicted to take ~350 sec to be synthesized, were subjected to a 45-sec  $^{35}\text{S}$ -methionine pulse. Labeling was terminated by arresting translation with cycloheximide, and UBP<sub>s</sub> were simultaneously inactivated with NEM (12). The release of modules **AUb** (DHFRhaUb), **BUb** ( $\text{Me}^K\beta\text{gal}$ Ub) and **CUb** ( $\text{Me}^K\text{DHFRha}$ Ub) by UBP cleavage was assayed by immunoprecipitation. **CUb\*** denotes both the  $\text{Me}^K\text{DHFRha}$ Ub band and a crossreacting band present in these NEM-treated extracts but not in untreated ones (compare with Fig. 3A).

**Figure 3.** A fraction of nascent polypeptides bearing an N-terminal degradation signal is degraded cotranslationally. (A) Determination of C/A ratios through immunoprecipitation of *in vivo*-labeled **AUb-BUb-CUb** fusion proteins (19). Two variants of  $\beta$ gal as domain **B** were used, one that carried an N-degron ( $\text{Re}^K\beta\text{gal}$ ,  $t_{1/2} \sim 2$  min) and the otherwise identical  $\beta$ gal that differed by one residue and lacked this degradation signal ( $\text{Me}^K\beta\text{gal}$ ,  $t_{1/2} > 20$  hr). The two  $\beta$ gal variants were expressed in *S. cerevisiae* strains containing different levels of Ubr1p, the rate-limiting recognition component of the N-end rule pathway. Each pair of lanes corresponds to two independent experiments. (B) Immunoblot analysis of the steady-state C/A levels for different **B** domains (described in Fig. 2A) (23). The identity of the N-terminal residue of each of the **B** domains, R (Arg) or M (Met), is indicated above each lane. F marks the lane corresponding to an **AUb-BUb-CUb** fusion protein in which domain **B** was the 18-residue FLAG-containing moiety.

**Figure 4.** The extent of cotranslational protein degradation depends on the presence of a degron, the activity of a degron-specific proteolytic pathway, and the nature and size of the protein. (A) C/A ratios obtained for the different **B** domains. (B) Ubr1p-dependent cotranslational degradation. These values were obtained by comparing C/A ratios observed for a particular N-degron-bearing **B** domain in wildtype or Ubr1p-overexpressing strains to the ratio determined in the *ubr1Δ* strain (20). Each bar represents a mean value derived from at least 4 independent experiments; standard errors are indicated. Differences between means were significant to at least  $p < 0.05$  by the Mann-Whitney test (30).

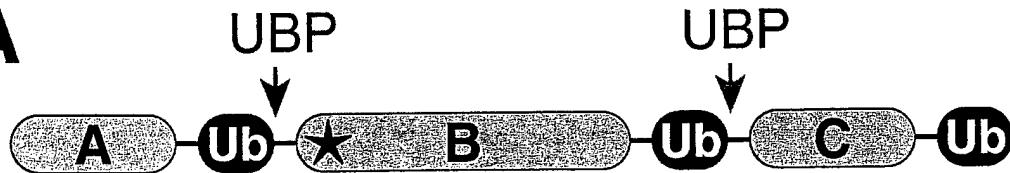
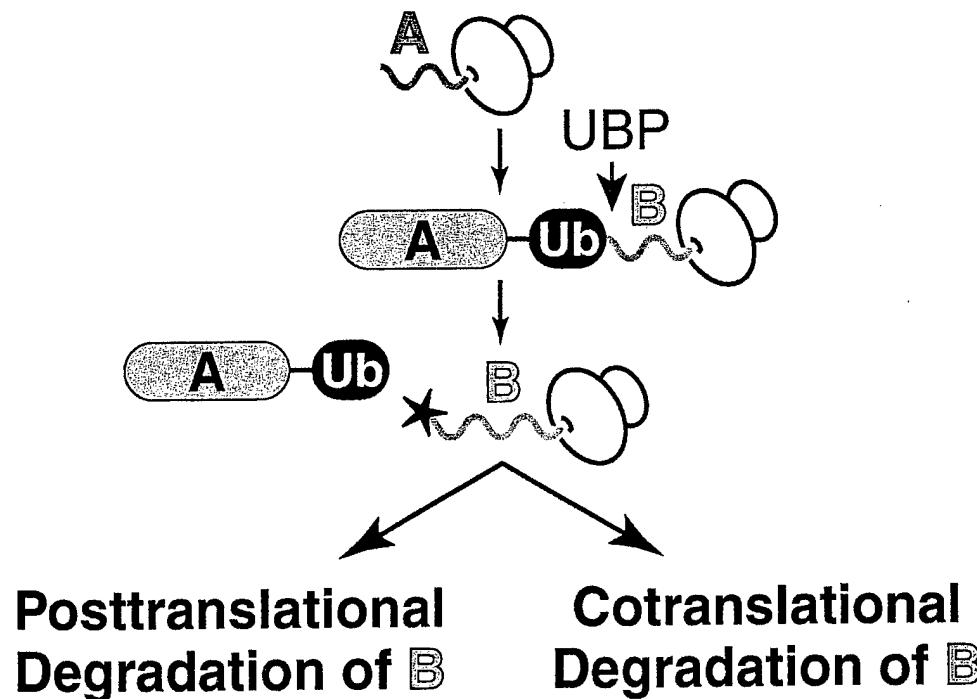
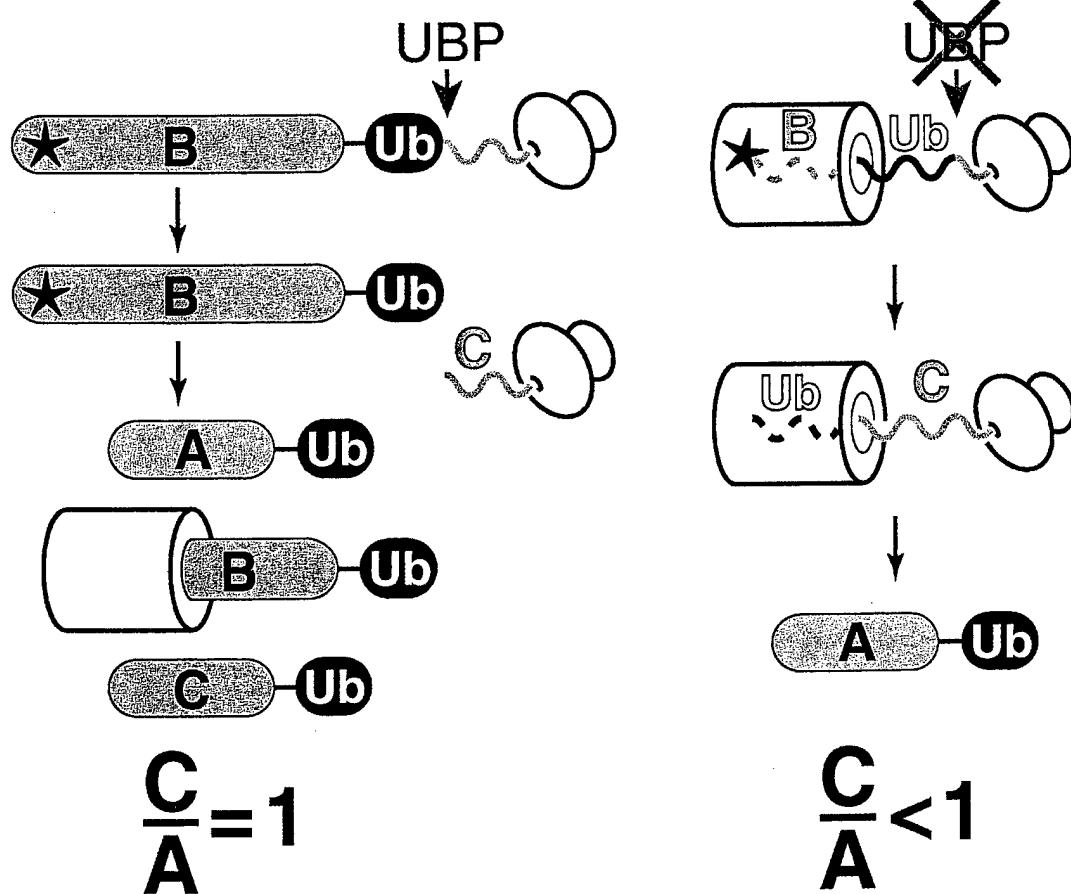
**A****B****Posttranslational  
Degradation of B****Cotranslational  
Degradation of B**

Figure 1

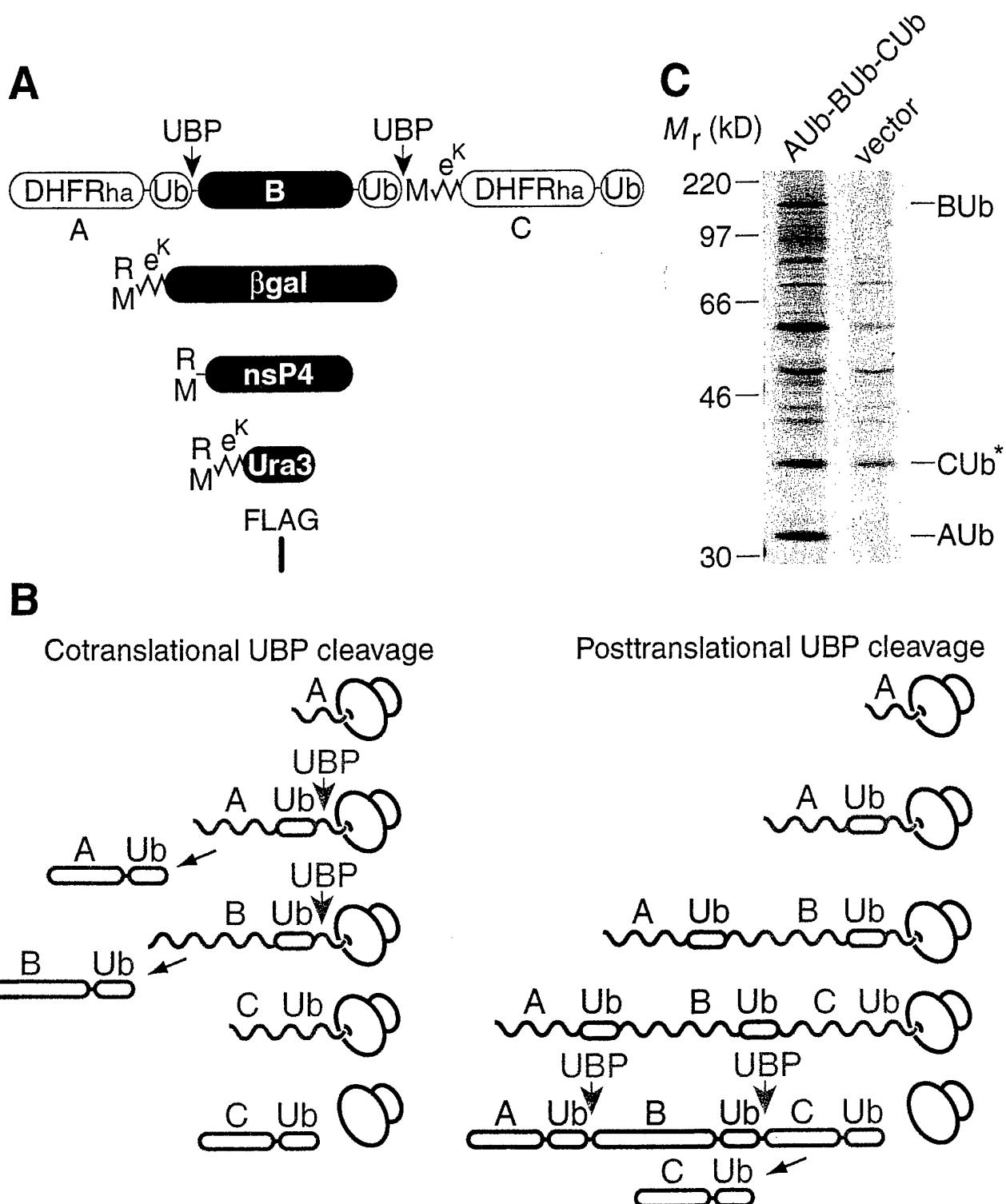


Figure 2

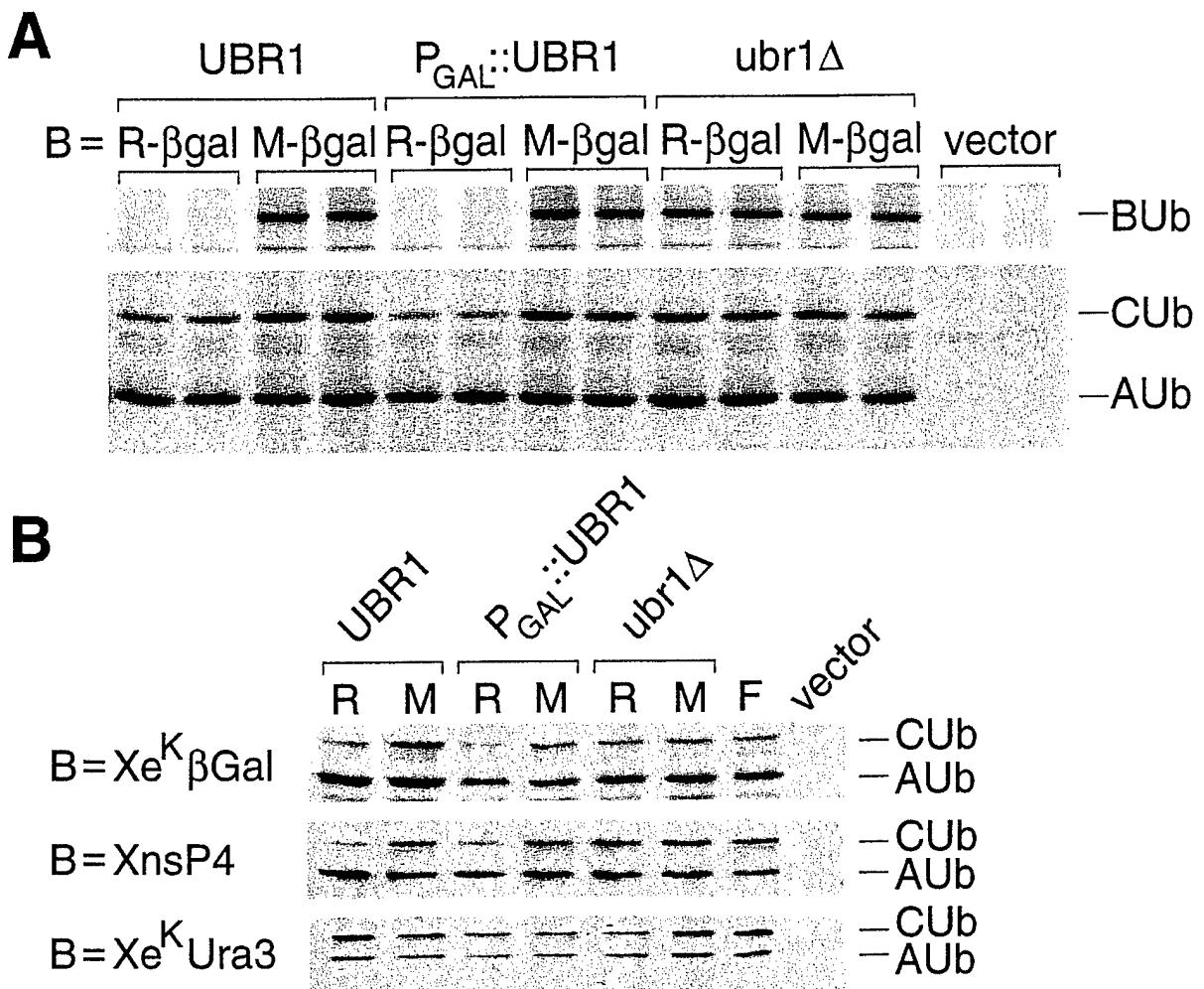


Figure 3

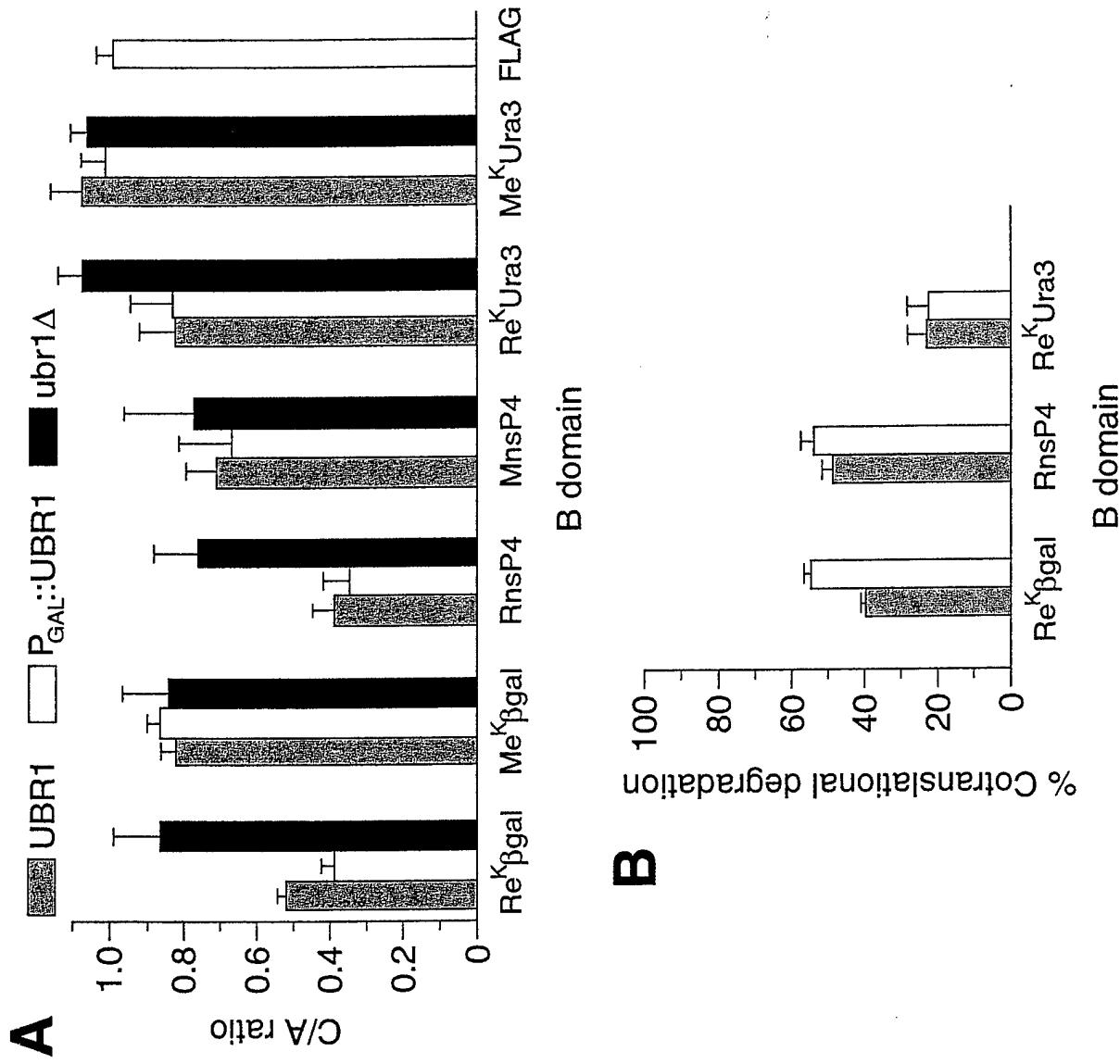


Figure 4

# Degradation signals in the lysine–asparagine sequence space

Tetsuro Suzuki<sup>1</sup> and Alexander Varshavsky<sup>2</sup>

Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

<sup>1</sup>Present address: Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

<sup>2</sup>Corresponding author  
e-mail:avarsh@its.caltech.edu

**The N-degrons, a set of degradation signals recognized by the N-end rule pathway, comprise a protein's destabilizing N-terminal residue and an internal lysine residue. We show that the strength of an N-degron can be markedly increased, without loss of specificity, through the addition of lysine residues. A nearly exhaustive screen was carried out for N-degrons in the lysine (K)–asparagine (N) sequence space of the 14-residue peptides containing either K or N (16 384 different sequences). Of these sequences, 68 were found to function as N-degrons, and three of them were at least as active and specific as any of the previously known N-degrons. All 68 K/N-based N-degrons lacked the lysine at position 2, and all three of the strongest N-degrons contained lysines at positions 3 and 15. The results support a model of the targeting mechanism in which the binding of the E3–E2 complex to the substrate's destabilizing N-terminal residue is followed by a stochastic search for a sterically suitable lysine residue. Our strategy of screening a small library that encompasses the entire sequence space of two amino acids should be of use in many settings, including studies of protein targeting and folding.**

**Keywords:** N-degron/N-end rule/proteolysis/simple sequences/ubiquitin

## Introduction

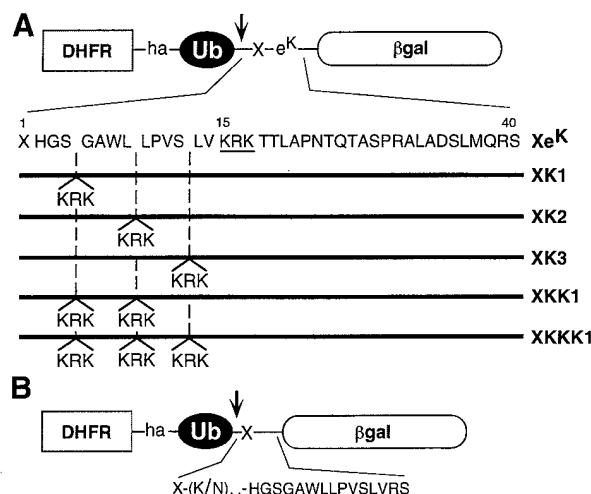
Regulatory proteins are often short-lived *in vivo*, providing a way to generate their spatial gradients and to rapidly adjust their concentration or subunit composition through changes in the rate of their synthesis or degradation. Most of the damaged or otherwise abnormal proteins are metabolically unstable as well. Many other proteins, while long-lived as components of larger structures such as ribosomes and oligomeric proteins, are short-lived as free subunits (reviewed by Hochstrasser, 1996; Varshavsky, 1997; Hershko and Ciechanover, 1998; Scheffner *et al.*, 1998; Koepp *et al.*, 1999; Tyers and Willems, 1999).

Features of proteins that confer metabolic instability are called degradation signals, or degrons (Laney and Hochstrasser, 1999). One class of degradation signals, called the N-degrons, comprises a protein's destabilizing

N-terminal residue and an internal Lys residue (Bachmair *et al.*, 1986; Varshavsky, 1996). A set of N-degrons containing different N-terminal residues that are destabilizing in a given cell defines a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The lysine determinant of an N-degron is the site of formation of a substrate-linked multi-ubiquitin chain (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (Pickart, 1997; Varshavsky, 1997; Peters, 1998; Scheffner *et al.*, 1998). In many of these settings, Ub acts through routes that involve the degradation of Ub–protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (Coux *et al.*, 1996; Hilt and Wolf, 1996; Baumeister *et al.*, 1998; Rechsteiner, 1998).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu (Baker and Varshavsky, 1995). The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase, to Arg, one of the primary destabilizing residues. The primary N-terminal residues are bound directly by the *UBR1*-encoded N-recognin, the E3 (recognition) component of the N-end rule pathway. In *S.cerevisiae*, N-recognin is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues: Phe, Leu, Trp, Tyr, Ile, Arg, Lys and His (Varshavsky, 1996). Analogous components of the mammalian N-end rule pathway have been identified as well (Stewart *et al.*, 1995; Grigoryev *et al.*, 1996; Kwon *et al.*, 1998, 1999).

Studies with engineered N-end rule substrates indicated the bipartite organization of N-degrons and suggested a stochastic model of their targeting, in which specific lysines of an N-end rule substrate could be assigned different probabilities of being used as a ubiquitylation site (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989; Johnson *et al.*, 1990; Hill *et al.*, 1993; Varshavsky, 1996; Lévy *et al.*, 1999). Most of the evidence for this model was produced with a set of N-degrons in which a destabilizing N-terminal residue X was linked to the ~40-residue *Escherichia coli* Lac repressor-derived sequence termed e<sup>K</sup> [extension (e) bearing lysines (K)] (Figure 1A) (Bachmair and Varshavsky, 1989). The resulting X-e<sup>K</sup> sequence comprised a portable N-degron, which could confer short half-lives on test proteins such as *E.coli* β-galactosidase



**Fig. 1.** Test proteins. (A) Fusions used in this work contained some of the following elements (see Materials and methods): DHFR-ha, a mouse dihydrofolate reductase moiety extended at the C-terminus by a sequence containing the hemagglutinin-derived ha epitope; the Ub<sup>R48</sup> moiety bearing the Lys→Arg alteration at position 48; a 40-residue *E.coli* Lac repressor-derived sequence, termed e<sup>K</sup> and shown in single-letter abbreviations for amino acids; a variable residue X (either Tyr, His or Met) between Ub<sup>R48</sup> and e<sup>K</sup>; the *E.coli* βgal moiety lacking the first 24 residues of wild-type βgal. The positions of single or multiple KRK insertions into e<sup>K</sup> are indicated. The endogenous KRK sequence of e<sup>K</sup> is underlined. The arrow indicates the site of *in vivo* cleavage by DUBs. (B) The fusion construct used for screening in the K/N sequence space. The 14 residues of e<sup>K</sup> immediately following the residue X were replaced by a set of 14-residue sequences that comprised a random permutation of Lys and Asn residues, followed by the sequence HGSGAWLLPVSLVRS, derived from residues 2–14 of the e<sup>K</sup> extension, followed by Arg-Ser.

(βgal) or mouse dihydrofolate reductase (DHFR) (Varshavsky, 1996). At least one of two lysines (K) in e<sup>K</sup>, either K-15 or K-17, must be present for the N-degron to be active (Figure 1A) (Bachmair and Varshavsky, 1989; Johnson *et al.*, 1990). Even though several other classes of N-degron, including the naturally occurring ones, have been described over the last decade (Townsend *et al.*, 1988; deGroot *et al.*, 1991; Dohmen *et al.*, 1994; Sadis and Finley, 1995; Ghislain *et al.*, 1996; Sijts *et al.*, 1997; Tobery and Siliciano, 1999), the mechanistic understanding of these degradation signals remains confined largely to the e<sup>K</sup>-based N-degrons (Varshavsky, 1996; Lévy *et al.*, 1999).

In the present work, we show that spiking an e<sup>K</sup>-based N-degron with additional Lys residues can markedly increase its activity. We also show, using a new approach of searching in the sequence space of lysine and asparagine, that simple-sequence N-degrons can be as strong and specific as any of the previously known N-degrons. These findings provide independent evidence for the model of a bipartite N-degron and stochastic targeting mechanism (Bachmair and Varshavsky, 1989). The strategy of exhaustive searching in the sequence space of two amino acids should be of use in many settings, including studies of protein folding and degradation.

## Results and discussion

### The Ub/protein/reference technique

The assays below utilized the previously developed Ub/protein/reference (UPR) technique, which increases the

accuracy of pulse-chase analysis by providing a ‘built-in’ reference protein (Lévy *et al.*, 1996). This method employs a linear fusion in which Ub is located between a protein of interest and a reference protein moiety (Figure 1A). The fusion is co-translationally cleaved by Ub-specific de-ubiquitylating enzymes (DUBs) (Wilkinson and Hochstrasser, 1998) after the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing the C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse-chase assay, the relative amounts of the protein of interest can be normalized against the reference protein in the same sample. The UPR technique can thus compensate for the scatter of immunoprecipitation yields, sample volumes and other sources of sample-to-sample variation (Lévy *et al.*, 1996, 1999).

Two previously introduced terms, ID<sup>x</sup>, initial decay, i.e. the decay of a protein during the pulse of  $x$  min, and  $t_{0.5}^{y-z}$ , the protein’s half-life averaged over the interval of  $y$  to  $z$  min of chase (Lévy *et al.*, 1996), are used below to describe the decay curves of test proteins. The ID<sup>x</sup> term and the interval-specific term  $t_{0.5}^{y-z}$  would be superfluous in the case of a strictly first-order decay, which is defined by a single half-life. However, the *in vivo* degradation of most proteins deviates from first-order kinetics. For example, the rate of degradation of short-lived proteins can be much higher during the pulse, in part because a newly labeled (either nascent or just-completed) polypeptide is conformationally immature and may, consequently, be targeted for degradation more efficiently than its mature counterpart. This enhanced early degradation, previously termed the ‘zero-point’ effect (Baker and Varshavsky, 1991), is described by the parameter ID<sup>x</sup> (Lévy *et al.*, 1996). It was found that a large fraction of the zero-point effect results from the co-translational degradation of nascent (being synthesized) polypeptide chains, which never reach their mature size before their destruction by processive proteolysis (G.Turner and A.Varshavsky, unpublished data). The detection of a zero-point effect requires the comparison of a test protein’s degradation between cells containing and lacking the relevant proteolytic pathway. Alternatively, the zero-point effect can be detected by comparing, through the UPR technique, the degradation of otherwise identical degron-containing and degron-lacking versions of a test protein (Lévy *et al.*, 1996, 1999). Although the degradation of a protein during the pulse can be strikingly high (Lévy *et al.*, 1996) (see also below), it is not detectable by a conventional, reference-lacking pulse-chase assay.

### Increasing the strength of N-degrons by spiking them with additional lysines

The UPR constructs of the present work were DHFR-ha-Ub<sup>R48</sup>-X-e<sup>K</sup>-βgal fusions. They contained the metabolically stable, ha-epitope-bearing DHFR-ha-Ub<sup>R48</sup> moiety as a reference protein, termed dha-Ub below. The dha-Ub-X-e<sup>K</sup>-βgal proteins were co-translationally cleaved *in vivo*, yielding the test protein X-e<sup>K</sup>-βgal and the reference dha-Ub (Figure 1A). To reduce the possibility that the C-terminal Ub moiety of dha-Ub could function as a ubiquitylation/degradation signal, the K-48 residue of Ub (a major site of isopeptide bonds in multi-Ub

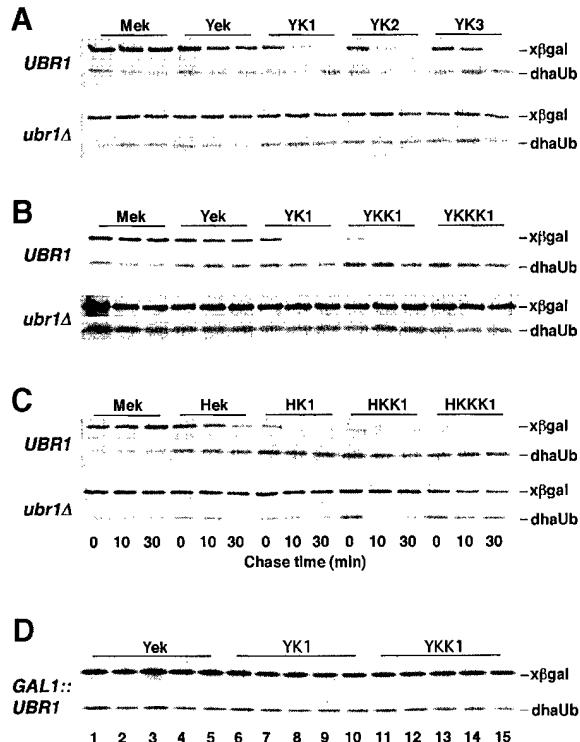
chains) was converted to Arg, which cannot be ubiquitylated, yielding Ub<sup>R48</sup> (Lévy *et al.*, 1996). These and related fusions (Figure 1A) were expressed in *S.cerevisiae* from low copy plasmids and the copper-inducible P<sub>CUP1</sub> promoter.

X-e<sup>K</sup>-βgal is an extensively analyzed class of N-end rule substrates, which contain a variable N-terminal residue X (produced through the DUB-mediated cleavage of dha-Ub-X-e<sup>K</sup>-βgal at the Ub-X junction), a 40-residue N-terminal extension called e<sup>K</sup> (see Introduction), and a βgal moiety lacking the first 24 residues of wild-type *E.coli* βgal (Figure 1A). If K-15 and K-17, the only lysines of the e<sup>K</sup> extension (Figure 1A), are replaced by Arg residues, which cannot be ubiquitylated, the resulting X-e<sup>AK</sup>-βgal is long-lived even if its N-terminal residue is destabilizing in the N-end rule (Bachmair *et al.*, 1986; Johnson *et al.*, 1990). The inactivity of N-degron in X-e<sup>AK</sup>-βgal is caused by the absence of targetable lysines (Varshavsky, 1996). Specifically, the multiple lysines of the βgal moiety in X-e<sup>AK</sup>-βgal (Chau *et al.*, 1989) cannot serve as N-degron determinants, apparently because the most N-terminal Lys residue in X-e<sup>AK</sup>-βgal, at position 239, is too far from the protein's N-terminus.

One of our aims was to produce stronger N-degrons. We chose Tyr, a moderately destabilizing type 2 residue (Bachmair and Varshavsky, 1989; Varshavsky, 1996), as the N-terminal residue of an initial test protein (Figure 1A). More strongly destabilizing N-terminal residues, e.g. Leu or Arg, in the context of (expected) stronger N-degrons would have made the test proteins too short-lived for detection in a pulse-chase assay. Met was employed as a stabilizing N-terminal residue. The term ID<sup>5</sup> below (see Materials and methods) conveys the extent of degradation of a protein during the 5 min pulse, in comparison with the degradation, during the same pulse, of a control (degron-lacking, i.e. Met-bearing) version of the same protein.

To determine whether the degradation of Tyr-e<sup>K</sup>-βgal in *S.cerevisiae* could be enhanced through the addition of Lys residues while remaining dependent on the Ubr1p N-recognin, the sequence Lys-Arg-Lys (KRK), identical to the sequence at positions 15–17 of e<sup>K</sup>, was inserted at the indicated locations within e<sup>K</sup> (Figure 1A). The unmodified Tyr-e<sup>K</sup>-βgal had an ID<sup>5</sup> of ~48%, i.e. ~48% of the labeled Tyr-e<sup>K</sup>-βgal was destroyed during the 5 min pulse, before time 0. The t<sub>0.5</sub><sup>0–10</sup> (half-life between 0 and 10 min of chase) of Tyr-e<sup>K</sup>-βgal was ~26 min (Figures 2A and 3A). The KRK sequence inserted at any of the indicated three positions within e<sup>K</sup> (Figure 1A) strongly destabilized the already short-lived Tyr-e<sup>K</sup>-βgal: for example, Tyr-K1e<sup>K</sup>-βgal (Figure 1A) had an ID<sup>5</sup> of ~75% and t<sub>0.5</sub><sup>0–10</sup> of ~5 min (Figures 2A and 3A). The increased degradation of Tyr-e<sup>K</sup>-βgal derivatives containing extra KRK remained completely Ubr1p-dependent: Tyr-e<sup>K</sup>-βgal, Tyr-K1e<sup>K</sup>-βgal, Tyr-K2e<sup>K</sup>-βgal and Tyr-K3e<sup>K</sup>-βgal were all long-lived proteins (t<sub>0.5</sub> > 10 h) in *ubr1Δ* cells (Figures 2A and 3A). In addition, Met-e<sup>K</sup>-βgal, Met-K1e<sup>K</sup>-βgal, Met-K2e<sup>K</sup>-βgal and Met-K3e<sup>K</sup>-βgal, the Met-bearing counterparts of the Tyr-based N-end rule substrates, were long-lived in either *UBR1* or *ubr1Δ* cells (data not shown).

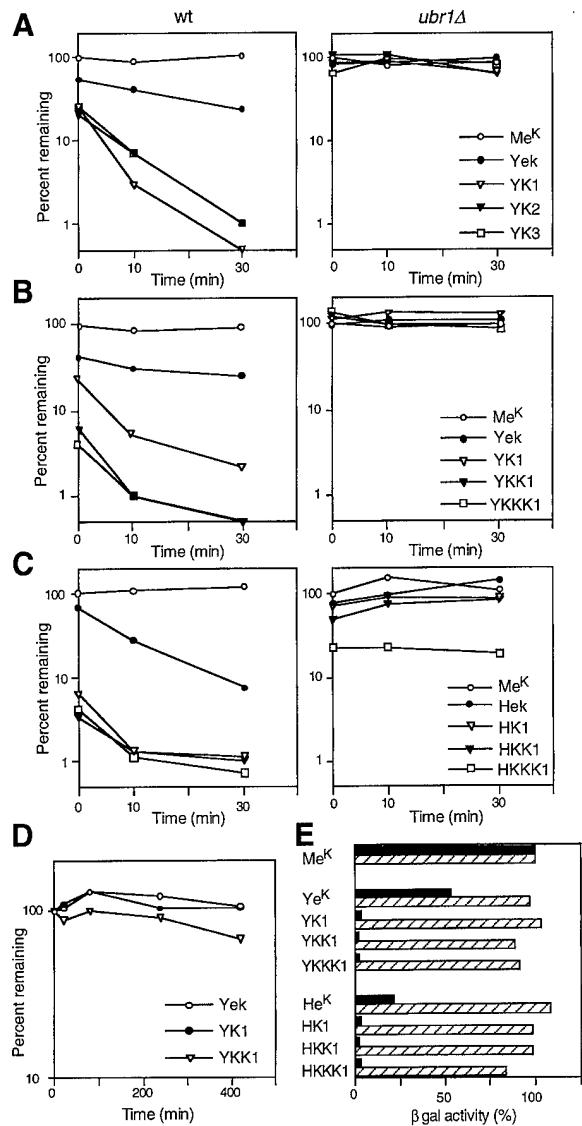
These results led us to examine the effects of adding more than one KRK sequence to e<sup>K</sup> (Figure 1A). The resulting Tyr-e<sup>K</sup>-βgal derivatives, bearing either two



**Fig. 2.** Active N-degrons can be strongly enhanced by additional lysines. (A) Congenic *UBR1* (wt) and *ubr1Δ* *S.cerevisiae* that expressed the UPR-based fusions Met-e<sup>K</sup>-βgal (Mek) (DHFR-ha-Ub<sup>R48</sup>-Met-e<sup>K</sup>-βgal), Tyr-e<sup>K</sup>-βgal (Yek), Tyr-K1e<sup>K</sup>-βgal (YK1), Tyr-K2e<sup>K</sup>-βgal (YK2) and Tyr-K3e<sup>K</sup>-βgal (YK3) (see Figure 1A) were labeled with [<sup>35</sup>S]methionine/cysteine for 5 min at 30°C, followed by a chase for 0, 10 and 30 min, extraction, immunoprecipitation with anti-ha and anti-βgal antibodies, SDS-PAGE, and autoradiography (see Materials and methods). The bands of X-βgal (test protein) and DHFR-ha-Ub<sup>R48</sup> (reference protein) are indicated on the right. (B) As in (A), but with Met-e<sup>K</sup>-βgal (Mek), Tyr-e<sup>K</sup>-βgal (Yek), Tyr-K1e<sup>K</sup>-βgal (YK1), Tyr-KK1e<sup>K</sup>-βgal (YKK1) and Tyr-KKK1e<sup>K</sup>-βgal (YKKK1) (see Figure 1A). (C) As in (A), but with Met-e<sup>K</sup>-βgal (Mek), His-e<sup>K</sup>-βgal (Hek), His-K1e<sup>K</sup>-βgal (HK1), His-KK1e<sup>K</sup>-βgal (HKK1) and His-KKK1e<sup>K</sup>-βgal (HKKK1). (D) Metabolic stability of conformationally mature Tyr-e<sup>K</sup>-βgal and its KRK-spiked derivatives. JD54 (P<sub>GAL1</sub>-UBR1) cells expressing Tyr-e<sup>K</sup>-βgal (Yek), Tyr-K1e<sup>K</sup>-βgal (YK1) or Tyr-KK1e<sup>K</sup>-βgal (YKK1) were grown in SM-raffinose medium (no expression of Ubr1p), then labeled with [<sup>35</sup>S]methionine/cysteine for 10 min at 30°C. After a 20 min chase in SM-raffinose, galactose was added to 3% to induce Ubr1p expression, followed by a chase for 1, 3 and 6 h, and the analysis of immunoprecipitated test proteins. Lanes 1, 6 and 11, the end of <sup>35</sup>S labeling (time 0). Lanes 2, 7 and 12, the end of 20 min chase in SM-raffinose. Lanes 3, 8 and 13, 1 h chase with galactose. Lanes 4, 9 and 14, 3 h chase. Lanes 5, 10 and 15, 6 h chase.

(Tyr-KK1e<sup>K</sup>-βgal) or three (Tyr-KKK1e<sup>K</sup>-βgal) KRK sequences, in addition to the original KRK of e<sup>K</sup>, were extremely short-lived proteins, even though N-terminal Tyr is a weakly destabilizing residue (Varshavsky, 1996). For example, Tyr-KK1e<sup>K</sup>-βgal (Figure 1A) had t<sub>0.5</sub><sup>0–10</sup> of ~4 min (in comparison with ~26 min in the case of Tyr-e<sup>K</sup>-βgal) and an ID<sup>5</sup> of ~94%. In other words, ~94% of the labeled Tyr-KK1e<sup>K</sup>-βgal was destroyed during the 5 min pulse, before time 0 (Figures 2B and 3B). At the same time, all of these proteins were long-lived in *ubr1Δ* cells (Figures 2B and 3B).

The N-terminal Tyr is bound by the type 2 site of N-recognin (Ubr1p) that recognizes substrates bearing bulky hydrophobic N-terminal residues (Varshavsky,



**Fig. 3.** Quantitation of degradation of the test proteins. (A) Pulse-chase patterns of Met-e<sup>K</sup>-βgal (Me<sup>K</sup>), Tyr-e<sup>K</sup>-βgal (Ye<sup>K</sup>), Tyr-K1e<sup>K</sup>-βgal (YK1), Tyr-K2e<sup>K</sup>-βgal (YK2) and Tyr-K3e<sup>K</sup>-βgal (YK3) (see Figures 1A and 2A) were quantitated using the UPR technique and PhosphorImager (see Materials and methods). Time 0 refers to the end of the 5 min pulse; 100% refers to the relative amount of Met-e<sup>K</sup>-βgal, normalized against the reference protein dha-Ub<sup>R48</sup>. (B) As in (A) but with Met-e<sup>K</sup>-βgal (Me<sup>K</sup>), Tyr-e<sup>K</sup>-βgal (Ye<sup>K</sup>), Tyr-K1e<sup>K</sup>-βgal (YK1), Tyr-KK1e<sup>K</sup>-βgal (YKK1) and Tyr-KKK1e<sup>K</sup>-βgal (YKKK1) (see Figures 1A and 2B). (C) As in (A) but with Met-e<sup>K</sup>-βgal (Me<sup>K</sup>), His-e<sup>K</sup>-βgal (He<sup>K</sup>), His-K1e<sup>K</sup>-βgal (HK1), His-KK1e<sup>K</sup>-βgal (HKK1) and His-KKK1e<sup>K</sup>-βgal (HKKK1) (see Figures 1A and 2C). (D) As in (A), but quantitation of the post-translational degradation of Tyr-e<sup>K</sup>-βgal (Ye<sup>K</sup>), Tyr-K1e<sup>K</sup>-βgal (YK1) and Tyr-KK1e<sup>K</sup>-βgal (YKK1) in the P<sub>GAL1</sub>-*UBR1* strain JD54 (see Figures 1A and 2A), following the induction of Ubr1p by galactose. Time 0 refers to the end of the 10 min labeling in raffinose (no Ubr1p). The cells were incubated for another 20 min in raffinose, followed by the addition of galactose to induce Ubr1p (see Materials and methods). (E) Relative enzymatic activities of βgal in *UBR1* cells (filled bars) and *ubr1Δ* cells (striped bars) expressing one of the following test proteins: Met-e<sup>K</sup>-βgal (Me<sup>K</sup>), Tyr-e<sup>K</sup>-βgal (Ye<sup>K</sup>), Tyr-K1e<sup>K</sup>-βgal (YK1), Tyr-KK1e<sup>K</sup>-βgal (YKK1), Tyr-KKK1e<sup>K</sup>-βgal (YKKK1), His-e<sup>K</sup>-βgal (He<sup>K</sup>), His-K1e<sup>K</sup>-βgal (HK1), His-KK1e<sup>K</sup>-βgal (HKK1) and His-KKK3e<sup>K</sup>-βgal (HKKK1) (see Materials and methods). The activities of βgal were normalized to the activity of Met-e<sup>K</sup>-βgal in each cell. Values shown are the means from duplicate measurements, which yielded results within 10% of the mean values.

1996). We asked whether the above findings were also relevant to the type 1 (basic) destabilizing N-terminal residues Arg, Lys and His, which are bound by the type 1 site of Ubr1p. Counterparts of the Tyr-βgal fusions that bore N-terminal His (a weak type 1 destabilizing residue) were constructed (Figure 1A) and tested in pulse-chase assays. The His residue was chosen as a type 1 destabilizing residue in these tests for the same reason as Tyr in the preceding tests: a stronger destabilizing residue would have made the measurements impractical with extremely short-lived substrates. The results (Figures 2C and 3C) confirmed the generality and specificity of degradation enhancement by the additional KRK sequences. For example, His-K1e<sup>K</sup>-βgal had an ID<sup>5</sup> of ~95% (i.e. ~95% of the labeled His-K1e<sup>K</sup>-βgal was destroyed during the 5 min pulse, before time 0), in comparison with the ID<sup>5</sup> of ~41% for His-e<sup>K</sup>-βgal; the corresponding  $t_{0.5}^{0-10}$  values were ~5 min and ~8 min for His-K1e<sup>K</sup>-βgal and His-e<sup>K</sup>-βgal, respectively (Figure 3C).

Similar to the results with Tyr-bearing substrates, their His-bearing, multiple KRK-containing counterparts were long-lived in *ubr1Δ* cells (Figures 2C and 3C). The only exception was His-KKK1e<sup>K</sup>-βgal (Figure 1A), which contained three KRK sequences, in addition to the KRK of the original e<sup>K</sup>: in contrast to Tyr-KKK1e<sup>K</sup>-βgal, His-KKK1e<sup>K</sup>-βgal was stabilized strongly but incompletely in the *ubr1Δ* genetic background (Figure 3C). Thus, the His-KKK1e<sup>K</sup> extension, in contrast to the Tyr-KKK1e<sup>K</sup> extension (Figure 1A), appears to contain a Ubr1p-independent degron.

Previous work (Madura *et al.*, 1993; Kwon *et al.*, 1999) has shown that the steady-state level of an X-βgal protein (determined by measuring the enzymatic activity of βgal in yeast extracts) is a sensitive measure of its metabolic stability. The results of this steady-state assay were in agreement with those derived from pulse-chase measurements: the level of Tyr-e<sup>K</sup>-βgal in *UBR1* cells was 53% of the level of the long-lived Met-e<sup>K</sup>-βgal, whereas Tyr-K1e<sup>K</sup>-βgal was present at 4% of the Met-e<sup>K</sup>-βgal level, and the concentration of Tyr-KKK1e<sup>K</sup>-βgal was virtually indistinguishable from the assay's background (cells transformed with vector alone) (Figure 3E). Crucially, the levels of these extra KRK-bearing Tyr-e<sup>K</sup>-βgal fusions in *ubr1Δ* cells became similar to that of Met-e<sup>K</sup>-βgal (Figure 3E), in agreement with the pulse-chase data (Figures 2A, B and 3A, B).

Although the addition of extra KRK sequences to e<sup>K</sup> yielded considerable decreases in the  $t_{0.5}^{0-10}$  of the corresponding N-end rule substrates, by far the major effect of multiple KRK sequences was on the decay curves' ID<sup>5</sup> term, which conveys the extent of degradation of a protein during or shortly after its synthesis (Figures 2A–C and 3A–C). To examine this issue in a different way, Tyr-e<sup>K</sup>-βgal, Tyr-K1e<sup>K</sup>-βgal and Tyr-KK1e<sup>K</sup>-βgal (Figure 1A) were produced in the JD54 *S.cerevisiae* strain, which expressed Ubr1p from the galactose-inducible, dextrose-repressible P<sub>GAL1</sub> promoter. JD54 cells expressing one of the test proteins were labeled in raffinose-containing SR medium (no Ubr1p), incubated for 20 min in the same medium and thereafter shifted to galactose, where Ubr1p was induced. Even though the N-end rule pathway became hyperactive in the presence of galactose (Madura and Varshavsky, 1994; Ghislain

*et al.*, 1996; data not shown), the pre-labeled substrates Tyr-e<sup>K</sup>-βgal and Tyr-K1e<sup>K</sup>-βgal were barely degraded after the induction of Ubr1p; Tyr-KK1e<sup>K</sup>-βgal was degraded only slightly (Figures 2D and 3D).

These findings were consistent with the earlier evidence for a strong retardation of the post-translational degradation of Arg-e<sup>K</sup>-βgal under the same conditions (R.J.Dohmen and A.Varshavsky, unpublished data). Thus, in contrast to a newly formed, conformationally immature βgal-based test protein, a conformationally mature βgal tetramer is a poor substrate of the N-end rule pathway even in the presence of N-degron enhancements such as the additional KRK sequences. It is the βgal moiety of these test proteins (Figure 1A) that was responsible for the time-dependent decline in the rate of degradation, because the kinetics of *in vivo* degradation of e<sup>K</sup>-DHFR-based N-end rule substrates was much closer to first-order decay (Lévy *et al.*, 1999; data not shown).

#### Locating N-degrons in the lysine-asparagine sequence space

The earlier work, which led to the bipartite model of N-degron (Bachmair and Varshavsky, 1989; Hill *et al.*, 1993), and particularly the present findings about the effects of adding KRK sequences to an e<sup>K</sup>-based N-degron (Figures 2 and 3) suggested that a substrate's destabilizing N-terminal residue and a sterically suitable internal Lys residue (or residues) are the two necessary and sufficient components of an N-degron. However, since both the e<sup>K</sup>-based and other previously analyzed N-degrons are embedded in complex sequence contexts (deGroot *et al.*, 1991; Dohmen *et al.*, 1994; Varshavsky, 1996), we wished to address the bipartite-degron model by constructing an N-degron from much simpler sequence motifs. Should this prove feasible, we also wanted to explore constraints on the structure of N-degrons through a screen in a simpler sequence setting. If the sequence space could be reduced strongly enough, one advantage of such a screen would be its exhaustiveness. The AAA codon for lysine differs by just one third-letter substitution from the codon for asparagine (AAU), a polar uncharged residue. Thus, one could attempt a screen for N-degrons in the sequence space of two amino acids: Lys (K) and Asn (N).

A double-stranded oligonucleotide that encoded random 14-residue K/N sequences (see Materials and methods) was used to replace the sequence encoding 14 residues of e<sup>K</sup> immediately following the residue X (Figure 1B). In the resulting test proteins, this latter sequence, HGSG-AWLLPVSLVRS (plus the sequence RS), followed the quasi-random 14-residue K/N sequence (Figure 1A). The resulting K/N-based extensions either lacked the lysines or contained a variable number of them between residues 2 and 16. The K-17 of e<sup>K</sup> (the only other lysine in e<sup>K</sup>) was replaced by Arg. In these test fusions, dha-Ub<sup>R48</sup>-Arg-(K/N)<sub>14</sub>-e<sup>Δ</sup>-βgals, Arg was used as a destabilizing N-terminal residue (Figure 1B). The number of different 14-residue sequences containing exclusively K or N is  $2^{14} = 16\,384$ . The bulk of a library of this complexity could be encompassed with conventional screening methods. Testing of the pRKN14-based library by amplifying it in *E.coli* indicated that >90% of the plasmids contained an oligonucleotide insert. The pRKN14 library was introduced into *S.cerevisiae* JD54 (Ghislain *et al.*, 1996), which

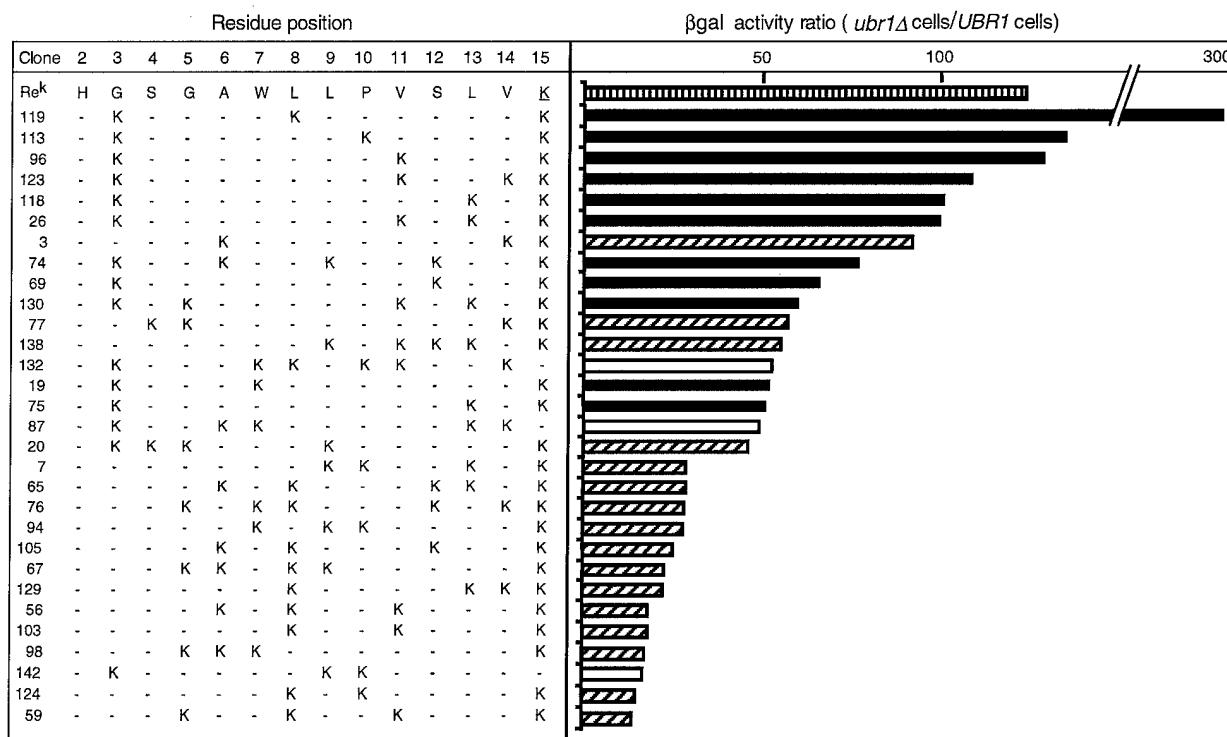
expressed Ubr1p from the *P<sub>GAL1</sub>* promoter, and screened for colonies that stained blue with XGal [high levels of Arg-(K/N)<sub>14</sub>-e<sup>Δ</sup>-βgal] on dextrose (SD) plates but stained white [low levels of Arg-(K/N)<sub>14</sub>-e<sup>Δ</sup>-βgal] on replica-plated galactose (SG) plates. Approximately 20 000 colonies were screened this way. A total of 68 isolates were identified in which the activity of βgal was significantly higher in the absence than in the presence of the N-end rule pathway.

The corresponding Arg-(K/N)<sub>14</sub>-e<sup>Δ</sup>-βgal test proteins were expressed in congenic *ubr1Δ* and *UBR1* strains, and the ratio of βgal activities was determined for each of the test proteins. The results are summarized in Figure 4, which shows the K/N sequences of the 30 most active N-degrons, and the ratios of the corresponding βgal activity in the *ubr1Δ* strain to that in the *UBR1* strain (higher ratios indicate stronger N-degrons). Remarkably, the strongest K/N-based N-degron was found to be more active than the strongest e<sup>K</sup>-based N-degron (Figure 4). Black bars in Figure 4 denote βgal activity derived from constructs carrying K/N-based N-degrons with lysines present at positions 3 and 15; the strongest N-degrons were largely of this class (Figure 4). K-15 was present in the 15 strongest K/N-based N-degrons except one (clone 132), which had K at position 14, and was also, presumably in compensation for the absence of K-15, one of the most lysine-rich N-degrons in this set (Figure 4). Similarly, K-3 was present in the 15 strongest K/N-based N-degrons except three (clones 3, 77 and 138) (Figure 4). All of these exceptional clones bore K-15; in addition, one of them (clone 77) bore K-4 and K-5, as well as K-14 and K-15 (Figure 4).

A completely uniform feature of all 68 K/N-based N-degrons was the absence of K from position 2 (Figure 4; data not shown), consistent with the fact that all of the previously examined N-degrons (Varshavsky, 1996) also lacked a strongly basic residue (Arg or Lys) at position 2. To address this issue directly, we used site-directed mutagenesis, replacing asparagine at position 2 of clone 119 (the strongest K/N-based N-degron; Figure 4) with lysine. The resulting test protein was long-lived in *UBR1* cells (data not shown), confirming that lysine is not tolerated at position 2 of an N-degron.

Even though all of the strongest K/N-based N-degrons contained at least three lysines, the total number of lysines in a K/N-based N-degron was not a strong predictor of its activity, and the arrangement of additional lysines between positions 3 and 15 did not correlate, in an obvious way, with the activity of N-degrons (Figure 4). None of the 15 strongest K/N-based N-degrons had more than six lysines (most had from three to five) (Figure 4), indicating that the density of lysines *per se* is not the main feature of a strong K/N-based N-degron. Note that our screen rejected K/N-based degrons that exhibited a significant Ubr1p-independent activity. This may in part account for the upper limit on the number of lysines in the strongest N-degrons: more lysine-rich K/N sequences could contain motifs recognized by non-N-end rule pathways, similarly to the above His-KKK1-βgal test protein (Figures 1A, 2C and 3C).

The *in vivo* degradation of Arg-(K/N)<sub>14</sub>-e<sup>Δ</sup>-βgal proteins bearing K/N-based N-degrons [produced from the dha-Ub<sup>R48</sup>-Arg-(K/N)<sub>14</sub>-e<sup>Δ</sup>-βgal fusions (Figure 1B)] was com-



**Fig. 4.** N-degrons in the K/N sequence space. The deduced sequences of the identified K/N N-degrons are shown in conjunction with the bar diagram of their relative activity, defined as the ratio of  $\beta$ -galactosidase (beta-gal) activities in the *ubr1* $\Delta$  versus *UBR1* cells expressing a given Arg-(K/N)<sub>14</sub>-e $^{\Delta}$ - $\beta$ -gal test protein. The top bar (vertical stripes) indicates the relative activity of N-degron in the original Arg-e $^K$ - $\beta$ -gal (Re $^K$ ). Thirty 14-residue K/N extensions with the highest Ubr1p-dependent destabilizing activity are listed, out of the total of 68 isolates that were metabolically unstable in the presence but not in the absence of Ubr1p. The total number of possible 14-residue K/N sequences is 16 384 (see the main text). The Lys and Asn residues are denoted as the letter K and a hyphen, respectively. Position 1 in each clone was occupied by Arg. The extensions in which lysines were present either at positions 3 and 15, or only at 15, or only at 3, are marked, respectively, by the filled, striped and open bars.

pared with the degradation of Arg-e $^K$ - $\beta$ -gal in pulse-chase assays (Figure 5A and B). The chosen K/N degrons (clones 119, 113 and 4) were the two most active ones in the  $\beta$ -gal assay (clones 119 and 113) (Figure 4), and a relatively weak one (clone 4) (Figure 6A). As expected from the results of steady-state  $\beta$ -gal assays (Figures 4), Arg-(K/N)<sub>14</sub>-e $^{\Delta}$ - $\beta$ -gal proteins 119 and 113 were extremely short-lived in *UBR1* cells (Figure 5A), in contrast to their stability in *ubr1* $\Delta$  cells (Figure 5B). As with Arg-e $^K$ - $\beta$ -gal, the bulk of degradation of Arg-(K/N)<sub>14</sub>-e $^{\Delta}$ - $\beta$ -gals took place either during or shortly after their synthesis, so that even at time 0 (at the end of the 5 min pulse) the test proteins could be detected only by overexposing the autoradiograms (Figure 5A; data not shown; compare with Figure 5B). In contrast, Arg-(K/N)<sub>14</sub>-e $^{\Delta}$ - $\beta$ -gal bearing the weaker N-degron of clone 4 was readily detectable at time 0, and decayed more slowly afterwards (Figure 5A), in agreement with the results of steady-state  $\beta$ -gal assays (Figure 6B). The distribution of three lysines in the strongest N-degron (clone 119: Lys3, 8, 15) (Figure 4) was similar to that in a relatively weak one (clone 4: Lys4, 12, 14) (Figure 6), emphasizing the importance of lysines at positions 3 and 15.

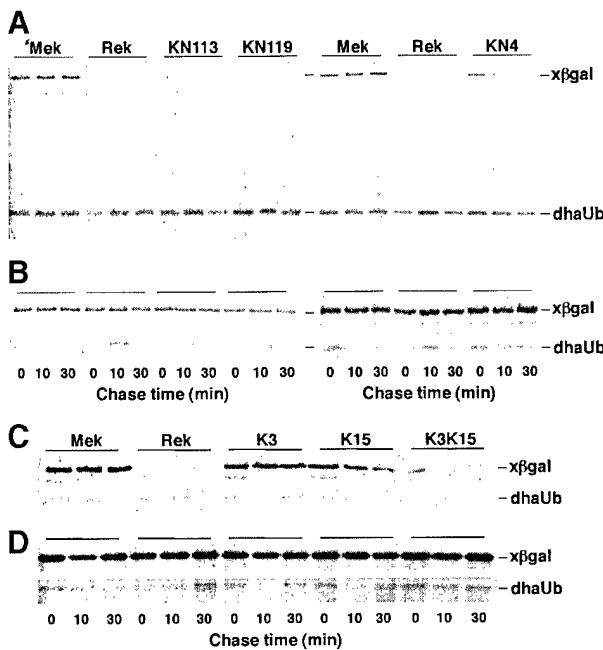
To address in more detail the relative contributions of the lysines at positions 3 and 15 to the activity of a K/N-based N-degron, site-directed mutagenesis was used to construct the otherwise identical N-degrons that contained either exclusively Lys3 and Lys15 (no lysine at a third, interior position), exclusively Lys3 or exclusively Lys15 (Figure 6A). The resulting Arg-(K/N)<sub>14</sub>-e $^{\Delta}$ - $\beta$ -gal

proteins were examined using both steady-state (Figure 6B) and pulse-chase assays (Figure 5C and D). The K/N-based N-degron that contained only Lys3 and Lys15 was active but considerably less so than the strongest K/N-based N-degron of clone 119, which contained Lys8 as well (Figures 4, 5C and 6). The elimination of either Lys3 or Lys15 further weakened the N-degron, so that the Lys3 and Lys15 versions of Arg-(K/N)<sub>14</sub>-e $^{\Delta}$ - $\beta$ -gal were, respectively, slightly and moderately short-lived proteins (Figures 5C, D and 7).

#### Mechanistic implications

One finding of this work is that an N-degron can be greatly strengthened by spiking it with additional lysine residues. The resulting degradation signals remained specific: nearly all of the enhanced N-degrons were completely inactive in *ubr1* $\Delta$  cells, which lacked the E3 (recognition) component of the N-end rule pathway. In addition, the enhanced N-degrons could be inactivated by replacing their destabilizing N-terminal residue with Met, a stabilizing residue in the N-end rule.

The strength of a K/N-based N-degron depends on the arrangement of lysines in the 14-residue N-terminal region of the test protein. This dependence is both strong and complex (Figures 4 and 6). The patterns observed may reflect distinct conformational flexibilities of different K/N sequences *vis-à-vis* the relatively fixed spatial arrangement of the type 1 or 2 sites of Ubr1p (N-recognition) and its associated Ubc2p E2 enzyme (Madura *et al.*, 1993).



**Fig. 5.** Pulse-chase analysis of K/N-based N-degrons. (A) JD47-13C (*UBR1*) *S.cerevisiae* that expressed the UPR-based fusions Met-e<sup>K</sup>-βgal (Mek) (DHFR-ha-Ub<sup>R48</sup>-Met-e<sup>K</sup>-βgal), R-e<sup>K</sup>-βgal (Rek), clone 113 of Arg-(K/N)<sub>14</sub>-e<sup>A</sup>-βgal (KN113), clone 119 of Arg-(K/N)<sub>14</sub>-e<sup>A</sup>-βgal (KN119), clone 4 of Arg-(K/N)<sub>4</sub>-e<sup>A</sup>-βgal (KN4) and other clones (see Figures 1B and 4) were labeled with [<sup>35</sup>S]methionine/cysteine for 5 min at 30°C, followed by a chase for 0, 10 and 30 min, extraction, immunoprecipitation with anti-ha and anti-βgal antibodies, SDS-PAGE and autoradiography (see Materials and methods). The bands of X-βgal (test protein) and DHFR-ha-Ub<sup>R48</sup> (reference protein) are indicated on the right. (B) As in (A), but with the congenic JD55 (*ubr1Δ*) cells. (C) As in (A) but with the clones K3, K15 and K3K15 (see Figure 6). (D) As in (C) but with JD55 (*ubr1Δ*) cells.

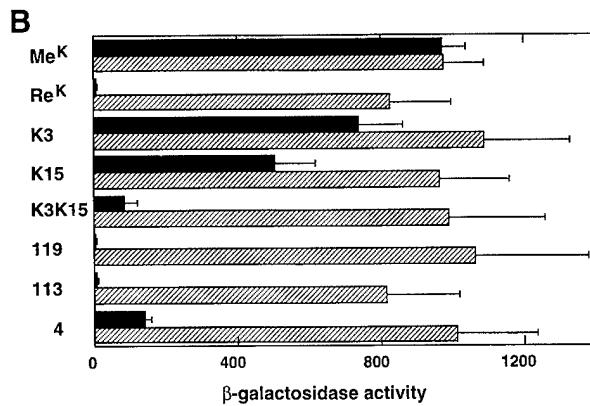
Some of the conclusions indicated by our data are described below.

(i) A completely uniform feature of all 68 K/N-based N-degrons was the absence of Lys from position 2 (Figure 4; data not shown), consistent with the fact that all of the previously examined N-degrons (Varshavsky, 1996) also lacked a strongly basic residue (Arg or Lys) at position 2. Using site-directed mutagenesis with clone 119 (Figure 4), we confirmed that lysine is not tolerated at position 2 of an N-degron (data not shown).

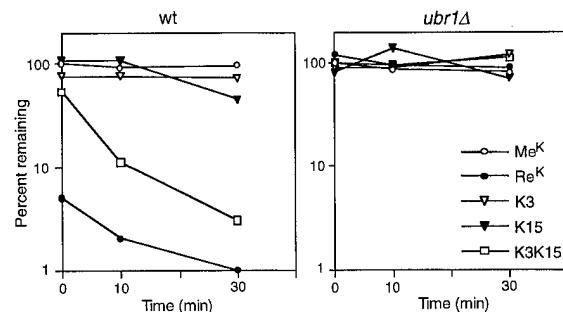
(ii) The strongest K/N-based N-degrons contained single Lys residues, surrounded by Asn residues (Figure 4), in contrast to the Lys-Arg-Lys (KRK) motif present in the e<sup>K</sup> extension of N-degrons studied previously (Figure 1A). We conclude that a lysine-containing motif of three adjacent basic residues is not an essential feature of an N-degron.

(iii) The presence of Lys15 in all of the most active K/N-based N-degrons (Figure 4) is consistent with the earlier model (Bachmair and Varshavsky, 1989; Varshavsky, 1996) in which a targetable Lys residue should be sufficiently far along the chain from a destabilizing N-terminal residue to allow the formation of a loop that positions this lysine spatially close to the N-terminal residue. This model is also consistent with the finding that Lys3 is another, nearly invariant, component of a K/N-based N-degron. Specifically, Lys3 may be located at the uniquely favorable ‘linear’ distance from the

Clone	Residue position															βgal activity ratio: <i>ubr1Δ</i> cells/ <i>UBR1</i> cells
	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
K3	-	K	-	-	-	-	-	-	-	-	-	-	-	-	-	1.5
K15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.9
K3K15	-	K	-	-	-	-	-	-	-	-	-	-	-	-	-	11.8
119	-	K	-	-	-	-	K	-	-	-	-	-	-	-	-	299.0
113	-	K	-	-	-	-	-	K	-	-	-	-	-	-	-	135.2
4	-	-	K	-	-	-	-	-	-	K	-	K	-	-	-	7.3



**Fig. 6.** Site-directed mutagenesis of N-degrons identified through the screen in the lysine-asparagine sequence space. (A) K/N sequences of clones 119, 113 and 4 and the K3/K15 derivatives of clone 119 (K3, K15 and K3K15). The relative activity of these N-degrons, defined as the ratio of βgal activities in the *ubr1Δ* versus *UBR1* cells expressing a given Arg-(K/N)<sub>14</sub>-e<sup>A</sup>-βgal protein, is indicated on the right. (B) The levels of βgal activity in extracts from *UBR1* cells (filled bars) and *ubr1Δ* cells (striped bars) cells expressing the indicated test proteins. Standard deviations (for triplicate measurements) are indicated.



**Fig. 7.** Decay curves of K3 and/or K15 mutants of Arg-(K/N)<sub>14</sub>-e<sup>A</sup>-βgal in *UBR1* and *ubr1Δ* *S.cerevisiae*. Pulse-chase patterns of Met-e<sup>K</sup>-βgal (MeK), Arg-e<sup>K</sup>-βgal (ReK), and Lys3 (K3), Lys15 (K15) and Lys3/Lys15 (K3K15) variants of Arg-(K/N)<sub>14</sub>-e<sup>A</sup>-βgal in Figure 5C and D were quantitated as described in the legend to Figure 3 and Materials and methods.

N-terminal residue, allowing its proximity to the destabilizing N-terminal residue in the absence of loop formation. That this view is at best incomplete is indicated by the finding that either Lys3 alone or Lys15 alone cannot substitute for the combination of Lys3, Lys8 and Lys15 that defines the strongest K/N-based N-degron identified to date (clone 119 in Figures 4 and 6). Moreover, even Lys3 and Lys15 together, in the absence of Lys8, result in a much weaker N-degron than the three-lysine N-degron of clone 119 (Figure 6). One possibility is that Lys15 is the only lysine, among the three, that can function as the site of ubiquitylation, and that lysines at positions 3 and 8 are required largely for optimal conformational flexibility

of the 14-residue K/N region in this N-degron. It is unlikely that the much higher activity of the clone 119 (three-lysine) N-degron resulted simply from its higher positive charge, in comparison with the two-lysine N-degrons, because most of the other K/N sequences containing three or more lysines were much less active than the clone 119 N-degron (Figure 4).

We searched for N-degrons in a K/N region 14 residues long (Figures 1B and 4), in part because the search with a significantly longer K/N region would have precluded the screen from being exhaustive. Nevertheless, since the e<sup>K</sup> extension of the earlier N-degrons is 40 residues long (Figure 1A), and since a part of the e<sup>K</sup> extension was retained, downstream of the K/N region, in our test proteins (see Materials and methods), it remains to be determined whether a K/N-based N-degron could be made even stronger by replacing the entire e<sup>K</sup> with a relevant K/N motif. For example, one could take advantage of the already identified 14-residue K/N sequence of clone 119 (Figure 4), and carry out an analogous screen for N-degrons in which the K/N sequence of clone 119 is fixed, while the downstream, e<sup>K</sup>-derived sequence is replaced by random K/N motifs. In sum, the K/N strategy of this work is far from exhausted by the present screen, and can be used, for example, to produce even stronger N-degrons, to search for other K/N-based degradation signals and to probe the targeting mechanisms of the Ub system.

### Concluding remarks

One implication of our results is the possibility of constructing much stronger N-degrons. These portable degradation signals can be used to render proteins of interest that are short-lived in either conditional or unconditional settings (Dohmen *et al.*, 1994; Worley *et al.*, 1998).

Another implication of our results stems from the demonstrated feasibility of defining a specific class of degradation signals in a sequence space of just two amino acids. Since the N-degrons have been defined previously in much more complex sequence contexts (Varshavsky, 1996), our findings suggest that other classes of degradation signal, and some of the other targeting signals as well, could also be identified and examined in this simple-sequence setting. Low-complexity sequences eliminate some of the informational ‘noise’ of natural sequences and thereby help to define the major determinants of structural specificity. The advantages of the simple-sequence approach have long been recognized by researchers who study the fundamentals of protein folding (Clarke, 1995). To our knowledge, the present work is the first to extend the simple-sequence approach to the realm of protein degradation.

## Materials and methods

### Strains, media and genetic techniques

*Saccharomyces cerevisiae* strains used in this work were JD47-13C (*MATa* *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112*) (Madura *et al.*, 1993), JD55 (*MATa* *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112 ubr1-Δ1::HIS3*) (Madura and Varshavsky, 1994) and JD54 (*MATa* *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112 GAL1::UBR1*) (Ghislain *et al.*, 1996). Rich (YPD) medium contained 1% yeast extract, 2% peptone (Difco) and 2% glucose. Synthetic media (Ausubel *et al.*, 1996) contained either 2% dextrose (glucose) (SD medium), 3% raffinose (SR medium) or 3% galactose (SG medium). To induce the P<sub>CUP1</sub> promoter,

CuSO<sub>4</sub> was added to a final concentration of 0.2 mM. Cells were incubated and assayed at 30°C. Transformation of *S.cerevisiae* was carried out using the lithium acetate method (Ausubel *et al.*, 1996).

### Plasmid construction

The plasmids encoded Ub fusions of the UPR technique (Lévy *et al.*, 1996) (see Results and discussion). The reference protein was mouse DHFR fused, through a 20-residue spacer containing the ha epitope, to the Ub<sup>R48</sup> moiety bearing Arg instead of wild-type Lys at position 48 (Lévy *et al.*, 1996). The plasmids pDhaUbXeKβgal expressed, from the P<sub>CUP1</sub> promoter, the fusions DHFR-ha-Ub<sup>R48-X-e<sup>K</sup>-βgal, where the junctional residue X was either Met, Tyr or His; e<sup>K</sup> was a previously described 40-residue, *E.coli* Lac repressor-derived N-terminal extension (Bachmair and Varshavsky, 1989; Johnson *et al.*, 1992); the *E.coli* βgal moiety lacked the first 24 residues of wild-type βgal (Bachmair *et al.*, 1986). The βgal-coding fragment was produced by PCR amplification of the βgal open reading frame (ORF) in pUB23 (Bachmair *et al.*, 1986; Bachmair and Varshavsky, 1989), using primers BGALS1 (5'-CAGAGATCTTTAATGCCCTTGACCA-3') and RSAS2 (5'-CCTCGAGGTCGACGGTATCG-3'), and the Expand PCR System (Boehringer, Indianapolis, IN). The Bg/II/Xhol-cut PCR product was ligated to the insert-lacking, P<sub>CUP1</sub> promoter-containing Bg/II-Xhol fragment of pdUbXekUra3, a pRS314-based, low copy plasmid (F.Lévy and A.Varshavsky, unpublished data). ORFs that expressed DHFR-ha-Ub<sup>R48-X-e<sup>K</sup>-βgal containing a modified e<sup>K</sup> region [single, double or triple insertions of the sequence Lys-Arg-Lys (KKR) (Figure 1A)] were produced using PCR. Specifically, pDhaUbXK1βgal, pDhaUbXK2βgal, pDhaUbXK3βgal, pDhaUbXKK1βgal and pDhaUbXKKK1βgal were constructed by replacing the BamHI-Bg/II fragment encoding X-e<sup>K</sup>-βgal with DNA fragments amplified using, respectively, the sense primers EKKS1 (5'-CACGGATCCAAGAGAAAAGGGAGCTTGGCTGTGC-CC-3'), EKKS2 (5'-CACGGATCCGGAGCTGGCTGAAGAGAAAAG-TTGGCCGTCTCACTGGTG-3'), EKKS3 (5'-CACGGATCCGGAG-CTTGGCTTGGCCGTCTCAAAGAGAAAAGCTGGTGAAAAGAA-AAACC-3'), EKKS4 (5'-CACGGATCCAAGAGAAAAGGGAGCTTGGCT-3') and EKKS1 (5'-CACGGATCCAAGAGAAAAGGGAGCTTGGCTGAAGAGAAAAG-TTGGCCGTCTCAAAGAGAAAAGCTGGTGAAAAGAAAAAAC-3'). The primer EKAS1 (5'-GGAAGATCTCTGCATTAATGAATC-3') and the plasmid pDhaUbMeKβgal served as the antisense primer and the template, respectively. The plasmids pDhaUbRK3eΔβgal, pDhaUbRK15-eΔβgal and pDhaUbRK3K15eΔβgal, which expressed, respectively, DHFR-ha-Ub<sup>R48</sup>-Arg-Asn-Lys-(Asn)<sub>12</sub>e<sup>A</sup>-βgal, DHFR-ha-Ub<sup>R48</sup>-Arg-(Asn)<sub>13</sub>-Lys-e<sup>A</sup>-βgal and DHFR-ha-Ub<sup>R48</sup>-Arg-Asn-Lys(Asn)<sub>11</sub>-Lys-e<sup>A</sup>-βgal, were constructed as follows. The term e<sup>A</sup> denotes the sequence HGSGAWLLPVSLVRS, a 13-residue derivative of the 40-residue e<sup>K</sup> (residues 2–14), followed by two residues Arg and Ser, encoded by the Bg/II site (Figure 1B). Double-stranded oligonucleotides that encompassed the 5' end of the SacII site, the end of Ub-Arg-(Lys/Asn)<sub>14</sub>-coding sequences (see below), and the 3' end of the BamHI site were digested with SacII and BamHI. The resulting fragments were ligated to the SacII-BamHI vector-containing fragment of pDhaUbReΔβgal. The latter plasmid was produced by inserting annealed complementary oligonucleotides EDKS1 and EDAS1, which encoded the e<sup>A</sup> sequence HGSGAWLLPVSLVRS (see above), into BamHI/Bg/II-cut pDhaUbReK-βgal. All constructs were verified by nucleotide sequencing.</sup></sup>

### A screen for N-degrons in the lysine-asparagine sequence space

The library of pDhaUbReΔβgal-derived plasmids was constructed that expressed, from the P<sub>CUP1</sub> promoter, Arg-(K/N)<sub>14</sub>e<sup>A</sup>-βgal proteins that contained random-sequence 14-residue Lys/Asn(K/N)-inserts between the N-terminal Arg and the e<sup>A</sup> moiety (Figure 1B). Arg-(K/N)<sub>14</sub>e<sup>A</sup>-βgals were the products of co-translational cleavage of DHFR-ha-Ub<sup>R48</sup>-Arg-(K/N)<sub>14</sub>e<sup>A</sup>-βgals (Figure 1B). To produce the library, a method for cloning random-sequence oligonucleotides produced by mutually primed synthesis was used (Oliphant *et al.*, 1986; Ghislain *et al.*, 1996). A set of oligonucleotides 5'-CGCCCGCGGTGGTAGG(AAA/T)<sub>14</sub>CACG-GATCCG-3' that contained random permutations of 14 codons, either AAA (encoding Lys) or AAT (encoding Asn), flanked by the SacII and BamHI sites (underlined), was synthesized. The oligonucleotides were converted into their double-stranded counterparts with Klenow Pol I (Ausubel *et al.*, 1996), and were digested with SacII and BamHI, yielding a set of equal-length fragments containing randomly permuted (AAA/T)<sub>14</sub> inserts. The fragments were ligated to the SacII-BamHI vector-containing fragment of pDhaUbReΔβgal. The resulting library, in the plasmid termed pRKN14, was introduced into *E.coli* DH5α by electropor-

ation (Ausubel *et al.*, 1996). Digestion of the pool of recovered plasmids with appropriate restriction enzymes showed that >90% of transformants contained an oligonucleotide-derived insert. The pRKN14 library was transformed into *S.cerevisiae* JD54, in which *Ubr1p* was expressed from the *P<sub>GAL1</sub>* promoter. Approximately  $2 \times 10^4$  transformants growing on SD(-Trp) plates containing 0.2 mM CuSO<sub>4</sub> were replica-plated onto CuSO<sub>4</sub>-containing SG(-Trp) plates. The XGal-based filter assay (Ausubel *et al.*, 1996) was used to screen for colonies that were white (low levels of βgal) on SG(-Trp) plates but blue (high levels of βgal) on SD(-Trp) plates. These colonies were grown up in liquid SG(-Trp), and the activity of βgal was determined as described below. Plasmid DNA was isolated from the positive transformants, amplified in *E.coli* and transformed into *S.cerevisiae* JD47-13C (*UBR1*) and JD55 (*ubr1Δ*). The metabolic stabilities of the corresponding Arg-(K/N)<sub>14</sub>-e<sup>A</sup>-βgal proteins in *ubr1Δ* versus wild-type (*UBR1*) *S.cerevisiae* were determined by measuring βgal activity, and also directly, by carrying out pulse-chase assays.

#### Measurement of βgal activity

*Saccharomyces cerevisiae* cells were added to 5 ml of SD(-Trp) containing 0.2 mM CuSO<sub>4</sub>, and grown to *A*<sub>600</sub> ~1. Cells were gently pelleted by centrifugation, lysed with glass beads in 20% glycerol, 1 mM dithiothreitol, 0.1 M Tris-HCl pH 8, and the activity of βgal was measured in the clarified extract using *o*-nitrophenyl-β-D-galactoside, as described (Ausubel *et al.*, 1996). The activity was normalized to the total protein concentration, determined using the Bradford assay (Bio-Rad, Hercules, CA).

#### Pulse-chase assays

Transformed JD47-13C and JD55 cells from 10 ml cultures (*A*<sub>600</sub> 0.5–1) in SD(-Trp) containing 0.2 mM CuSO<sub>4</sub> were gently pelleted by centrifugation, and washed in the same medium. The cells were resuspended in 0.3 ml of SD(-Trp) containing CuSO<sub>4</sub> and labeled for 5 min at 30°C with 0.15 mCi (5.5 MBq) of [<sup>35</sup>S]methionine/cysteine (Trans-<sup>35</sup>S-label, ICN, Costa Mesa, CA). The cells were harvested by centrifugation, resuspended in 0.3 ml of SD(-Trp), 10 mM L-methionine, 0.5 mg/ml cycloheximide and incubated further at 30°C. At each time point, 0.1 ml samples were withdrawn and added to 0.7 ml of the lysis buffer (1% Triton X-100, 0.15 M NaCl, 1 mM EDTA, 50 mM Na-HEPES pH 7.5) containing 1 mM phenylmethylsulfonyl fluoride. The cells were then lysed by vortexing with 0.5 ml of 0.5 mm glass beads four times for 1 min, with intermittent cooling on ice, followed by centrifugation at 12 000 g for 10 min. The volumes of supernatants were adjusted to equalize the amounts of 10% trichloroacetic acid-insoluble <sup>35</sup>S, followed by immunoprecipitation with a mixture of saturating amounts of monoclonal antibodies against the ha epitope (Babco, Berkeley, CA) and βgal (Promega, Madison, WI). The samples were incubated at 4°C for 2 h, with rotation, followed by the addition of 10 µl of protein A-Sepharose suspension (Repligen, Cambridge, MA), further incubation for 1 h, and centrifugation at 12 000 g for 30 s. The immunoprecipitates were washed four times with 0.8 ml of the lysis buffer plus 0.1% SDS, resuspended in SDS-sample buffer (Ausubel *et al.*, 1996), heated at 100°C for 3 min and fractionated by SDS-10% PAGE, followed by autoradiography and quantitation with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA), using the reference provided by the UPR technique (Figures 2 and 5) (Lévy *et al.*, 1996).

Pulse-chase assays with JD54 cells, in which *UBR1* was expressed from the *P<sub>GAL1</sub>* promoter, were carried out by growing 20 ml cultures to *A*<sub>600</sub> 0.5–1 in SR(-Trp), collecting and resuspending the cells in 0.5 ml of the same medium containing 0.2 mM CuSO<sub>4</sub> and labeling with 0.3 mCi of [<sup>35</sup>S]methionine/cysteine for 10 min at 30°C. The cells were washed twice with SR(-Trp), and transferred to 0.5 ml of SR (-Trp) containing 10 mM L-methionine and 1 mM cysteine. After a further 20 min incubation at 30°C, galactose was added (to a final concentration of 3%) to induce the expression of *UBR1* and initiate the chase, whose time points are indicated in Figures 2D and 3D.

ID<sup>5</sup>, the extent of initial decay at the end of the 5 min pulse, was calculated as follows: ID<sup>5</sup> = {1 - [Xβgal]<sub>0</sub>/[Met-e<sup>K</sup>βgal]<sub>0</sub>} × 100%. This parameter (the upper index refers to the length of pulse) equals 100% minus the ratio of <sup>35</sup>S in an X-βgal to <sup>35</sup>S in the reference protein dha-Ub (DHFR-ha-Ub<sup>R48</sup>) at the end of the pulse, normalized against the same ratio with metabolically stable Met-e<sup>K</sup>-βgal (Lévy *et al.*, 1996). To denote the observed half-lives of a test protein at different regions of a non-exponential decay curve, a generalized half-life term  $t_{\frac{y-z}{z}}$  was used, in which 0.5 denotes the parameter's half-life aspect and y-z denotes the relevant time interval, from y to z min of chase.

#### Acknowledgements

We thank the current and former members of the Varshavsky laboratory, particularly F.Lévy, A.Webster and Y.Xie, for helpful discussions. We also thank F.Du, J.Sheng, H.-R.Wang, Y.Xie, H.Rao and especially G.Turner for comments on the manuscript. This work was supported by grants to A.V. from the National Institutes of Health (DK39520) and the US Army Breast Cancer Research Program (DAMD179818042).

#### References

- Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (eds) (1996) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York, NY.
- Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell*, **56**, 1019–1032.
- Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.
- Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.
- Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.
- Baumeister,W., Walz,J., Zühl,F. and Seemüller,E. (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **92**, 367–380.
- Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.
- Clarke,N.D. (1995) Sequence ‘minimization’: exploring the sequence landscape with simplified sequences. *Curr. Opin. Biotechnol.*, **6**, 467–472.
- Coux,O., Tanaka,K. and Goldberg,A.L. (1996) Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.*, **65**, 801–817.
- deGroot,R.J., Rümenapf,T., Kuhn,R.J. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.
- Dohmen,R.J., Wu,P. and Varshavsky,A. (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science*, **263**, 1273–1276.
- Ghislain,M., Dohmen,R.J., Lévy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *EMBO J.*, **15**, 4884–4899.
- Grigoryev,S., Stewart,A.E., Kwon,Y.T., Arfin,S.M., Bradshaw,R.A., Jenkins,N.A., Copeland,N.G. and Varshavsky,A. (1996) A mouse amidase specific for N-terminal asparagine. The gene, the enzyme and their function in the N-end rule pathway. *J. Biol. Chem.*, **271**, 28521–28532.
- Hershko,A. and Ciechanover,A. (1998) The ubiquitin system. *Annu. Rev. Biochem.*, **76**, 425–479.
- Hill,C.P., Johnston,N.L. and Cohen,R.E. (1993) Crystal structure of a ubiquitin dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl Acad. Sci. USA*, **90**, 4136–4140.
- Hilt,W. and Wolf,D.H. (1996) Proteasomes: destruction as a programme. *Trends Biochem. Sci.*, **21**, 96–102.
- Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.
- Johnson,E.S., Gonda,D.K. and Varshavsky,A. (1990) *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. *Nature*, **346**, 287–291.
- Johnson,E.S., Bartel,B.W. and Varshavsky,A. (1992) Ubiquitin as a degradation signal. *EMBO J.*, **11**, 497–505.
- Koeppl,D.M., Harper,J.W. and Elledge,S.J. (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell*, **97**, 431–434.
- Kwon,Y.T. *et al.* (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **95**, 7898–7903.
- Kwon,Y.T., Kashina,A.S. and Varshavsky,A. (1999) Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. *Mol. Cell. Biol.*, **19**, 182–193.
- Laney,J.D. and Hochstrasser,M. (1999) Substrate targeting in the ubiquitin system. *Cell*, **97**, 427–430.
- Lévy,F., Johnsson,N., Rumenapf,T. and Varshavsky,A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl Acad. Sci. USA*, **93**, 4907–4912.

- Lévy,F., Johnston,J.A. and Varshavsky,A. (1999) Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur. J. Biochem.*, **259**, 244–252.
- Madura,K. and Varshavsky,A. (1994) Degradation of Ga by the N-end rule pathway. *Science*, **265**, 1454–1458.
- Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.
- Oliphant,A.R., Nussbaum,A.L. and Struhl,K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene*, **44**, 177–183.
- Peters,J.M. (1998) SCF and APC: the Yin and Yang of cell cycle regulated proteolysis. *Curr. Opin. Cell Biol.*, **10**, 759–768.
- Pickart,C.M. (1997) Targeting of substrates to the 26S proteasome. *FASEB J.*, **11**, 1055–1066.
- Rechsteiner,M. (1998) The 26S proteasome. In Peters,J.M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 147–189.
- Sadis,S.C.A. and Finley,D. (1995) Synthetic signals for ubiquitin-dependent proteolysis. *Mol. Cell. Biol.*, **15**, 4086–4094.
- Scheffner,M., Smith,S. and Jentsch,S. (1998) The ubiquitin conjugation system. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 65–98.
- Sijts,A.J., Pilip,I. and Pamer,E.G. (1997) The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class I antigen processing of bacterial proteins. *J. Biol. Chem.*, **272**, 19261–19268.
- Stewart,A.E., Arfin,S.M. and Bradshaw,R.A. (1995) The sequence of porcine protein NH<sub>2</sub>-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 25–28.
- Tobery,T. and Siliciano,R.F. (1999) Induction of enhanced CTL-dependent protective immunity *in vivo* by N-end rule targeting of a model tumor antigen. *J. Immunol.*, **162**, 639–642.
- Townsend,A., Bastin,J., Gould,K., Brownlee,G., Andrew,M., Coupar,B., Boyle,D., Chan,S. and Smith,G. (1988) Defective presentation to class I-restricted cytotoxic T lymphocytes in vaccinia-infected cells is overcome by enhanced degradation of antigen. *J. Exp. Med.*, **168**, 1211–1224.
- Tyers,M. and Willems,A.R. (1999) One ring to rule a superfamily of E3 ubiquitin ligases. *Science*, **284**, 602–604.
- Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.
- Varshavsky,A. (1997) The ubiquitin system. *Trends Biochem. Sci.*, **22**, 383–387.
- Wilkinson,K. and Hochstrasser,M. (1998) The deubiquitinating enzymes. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY.
- Worley,C.K., Ling,R. and Callis,J. (1998) Engineering *in vivo* instability of firefly luciferase and *Escherichia coli* beta-glucuronidase in higher plants using recognition elements from the ubiquitin pathway. *Plant Mol. Biol.*, **37**, 337–347.

*Received August 18, 1999; revised September 15, 1999;  
accepted September 16, 1999*

## Alternative Splicing Results in Differential Expression, Activity, and Localization of the Two Forms of Arginyl-tRNA-Protein Transferase, a Component of the N-End Rule Pathway

YONG TAE KWON, ANNA S. KASHINA, AND ALEXANDER VARSHAVSKY\*

Division of Biology, California Institute of Technology, Pasadena, California 91125

Received 13 August 1998/Returned for modification 21 September 1998/Accepted 6 October 1998

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The underlying ubiquitin-dependent proteolytic system, called the N-end rule pathway, is organized hierarchically: N-terminal aspartate and glutamate (and also cysteine in metazoans) are secondary destabilizing residues, in that they function through their conjugation, by arginyl-tRNA-protein transferase (R-transferase), to arginine, a primary destabilizing residue. We isolated cDNA encoding the 516-residue mouse R-transferase, ATE1p, and found two species, termed *Ate1-1* and *Ate1-2*. The *Ate1* mRNAs are produced through a most unusual alternative splicing that retains one or the other of the two homologous 129-bp exons, which are adjacent in the mouse *Ate1* gene. Human *ATE1* also contains the alternative 129-bp exons, whereas the plant (*Arabidopsis thaliana*) and fly (*Drosophila melanogaster*) *Ate1* genes encode a single form of ATE1p. A fusion of ATE1-1p with green fluorescent protein (GFP) is present in both the nucleus and the cytosol, whereas ATE1-2p-GFP is exclusively cytosolic. Mouse ATE1-1p and ATE1-2p were examined by expressing them in *ate1Δ Saccharomyces cerevisiae* in the presence of test substrates that included Asp-βgal (β-galactosidase) and Cys-βgal. Both forms of the mouse R-transferase conferred instability on Asp-βgal (but not on Cys-βgal) through the arginylation of its N-terminal Asp, the ATE1-1p enzyme being more active than ATE1-2p. The ratio of *Ate1-1* to *Ate1-2* mRNA varies greatly among the mouse tissues; it is ~0.1 in the skeletal muscle, ~0.25 in the spleen, ~3.3 in the liver and brain, and ~10 in the testis, suggesting that the two R-transferases are functionally distinct.

The half-lives of intracellular proteins range from a few seconds to many days. The rates of processive proteolysis are a function of the cell's physiological state and are controlled differentially for specific proteins. In particular, most of the damaged or otherwise abnormal proteins are metabolically unstable. Many other proteins, while long-lived as components of larger macromolecular structures such as ribosomes and oligomeric proteins, are metabolically unstable as free subunits. Regulatory proteins are often also short-lived *in vivo*, providing a way to generate their spatial gradients and to rapidly adjust their concentrations, or subunit compositions, through changes in the rate of their synthesis or degradation (20, 23, 28, 39, 44, 55).

The posttranslational conjugation of arginine (Arg) to the N termini of eukaryotic proteins was described 35 years ago (26), but the function of this modification, and of the enzyme involved, Arg-tRNA-protein transferase (R-transferase) (47), remained unknown until the discovery that the identity of N-terminal residue in a protein influences its metabolic stability (4). The resulting relation was termed the N-end rule (54). Aspartate (Asp) and glutamate (Glu), the two N-terminal residues known to be arginylated by R-transferase (47), were shown to be destabilizing residues in the N-end rule (4). It was therefore proposed (4) that the function of R-transferase is to target proteins for degradation by conjugating Arg, one of the primary destabilizing residues, to secondary destabilizing N-terminal residues (Asp and Glu in fungi; Asp, Glu, and Cys in metazoans) (18) (Fig. 1). It was also proposed (4) that the analogous prokaryotic enzyme Leu, Phe-tRNA-protein transferase (L, F-transferase) (47) mediates the activity of N-termini

nal Arg and Lys, which, in prokaryotes, would be the secondary destabilizing residues. These conjectures were confirmed (7, 17, 45, 53).

The similar but distinct degradation signals which together give rise to the N-end rule are called the N-degrons (54, 56). In eukaryotes, an N-degron comprises two determinants: a destabilizing N-terminal residue and an internal Lys residue of a substrate (5, 22). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (11). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (20, 23, 39, 55). In many of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (9, 13, 40, 43).

(In the text that follows, names of mouse genes are in italics, with the first letter uppercase. Names of human and *Saccharomyces cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. Names of *S. cerevisiae* proteins are roman, with the first letter uppercase and an extra lowercase "p" at the end. Names of the corresponding mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention (50), to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.)

The N-end rule is organized hierarchically. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase (Nt-amidase) (6), to the secondary destabilizing N-terminal residues Asp and Glu. The

\* Corresponding author. Mailing address: Division of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125. Phone: (626) 395-3785. Fax: (626) 440-9821. E-mail: avarsh@cco.caltech.edu.

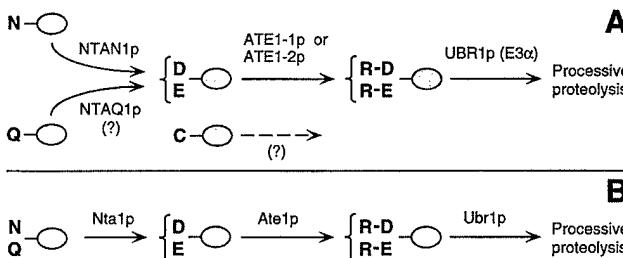


FIG. 1. Comparison of enzymatic reactions that underlie the activity of the tertiary and secondary destabilizing residues among eukaryotes. (A) Mammals (reference 54 and this work); (B) the yeast *S. cerevisiae* (5). N-terminal residues are indicated by single-letter abbreviations for amino acids; ovals denote the rest of a protein substrate. The *Ntan1*-encoded mammalian Nt-amidase converts N-terminal Asn to Asp, whereas N-terminal Gln is deamidated by a distinct Nt-amidase that remains to be identified (19, 51). In contrast, the yeast Nt-amidase can deamidate either N-terminal Asn or Gln (6). The secondary destabilizing residues Asp and Glu are arginylated by the mammalian ATE1-1p or ATE1-2p R-transferase (see Results). A Cys-specific mammalian R-transferase remains to be identified (see Results). N-terminal Arg, one of the primary destabilizing residues (54), is recognized by N-recognin (E3) (see the introduction).

destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *S. cerevisiae* *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), to Arg, one of the primary destabilizing residues (7) (Fig. 1B). In mammals, the deamidation step is bifurcated, in that two distinct Nt-amidases specific, respectively, for N-terminal Asn and Gln, mediate the activity of tertiary destabilizing residues (19, 51) (Fig. 1A). Mice lacking the Asn-specific Nt-amidase NTAN1p have recently been produced through targeted mutagenesis and found to be fertile, outwardly normal, but behaviorally distinct from their congenic wild-type counterparts (28a). In mammals, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (18, 54) (Fig. 1).

The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3 $\alpha$ ), the recognition component of the N-end rule pathway (8). In *S. cerevisiae*, N-recognin is a 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues, Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (54). The *Ubr1* genes encoding mouse and human N-recognins, also called E $\alpha$  (21, 41), have been cloned (29), and mouse strains lacking *Ubr1* have recently been constructed (29a).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2* which encodes the peptide transporter (2, 10); a role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (38); the degradation of Gpa1p, a G $\alpha$  protein of *S. cerevisiae* (34); and the conditional degradation of alphaviral RNA polymerase in virus-infected metazoan cells (16, 54). Physiological N-end rule substrates were also identified among the proteins secreted into the host cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes* (46). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (24) and to delay limb regeneration in amphibians (52). Microarray-based comparisons of gene expression patterns in wild-type and congenic *ubr1* $\Delta$  strains of *S. cerevisiae* have shown that a number of yeast genes, of diverse functions, are significantly

up- or down-regulated in the absence of the N-end rule pathway (39a).

The mammalian counterpart of the yeast *ATE1*-encoded R-transferase was partially purified from rabbit reticulocytes and shown to cofractionate with Arg-tRNA synthetase (12). Recent studies of the Ub-dependent proteolysis of endogenous proteins in muscle extracts suggested that the N-end rule pathway plays a major role in catabolic states that result in muscle atrophy (48, 49). A significant fraction of the N-end rule pathway's activity in muscle extracts was found to be tRNA dependent, indicating the involvement of R-transferase (48, 49). It was also reported that a crush injury to the rat sciatic nerve results in a ~10-fold increase in the rate of arginine conjugation to the N termini of unidentified proteins in the nerve's region upstream of the crush site (15, 57), suggesting an injury-induced increase in the concentration of R-transferase substrates and/or an enhanced activity of the N-end rule pathway.

In this work, we began the functional analysis of mammalian R-transferase (ATE1p) by isolating mouse cDNA encoding this enzyme. Surprisingly, we found two *Ate1* cDNA species, which were identical except for a 129-bp region that encoded similar but distinct sequences. One or the other, but not both, of the corresponding *Ate1* exons is retained in the mature *Ate1* mRNA, and the ratio of the resulting two species, *Ate1-1* and *Ate1-2*, varies greatly among mouse tissues. We also show that ATE1-1p and ATE1-2p, while differing in activity, can arginylate N-terminal Asp and Glu in model substrates. However, neither of them can arginylate N-terminal Cys, the known secondary destabilizing residue (18), suggesting the existence of a distinct Cys-specific mammalian R-transferase.

#### MATERIALS AND METHODS

**Strains and plasmids.** The *S. cerevisiae* strains used were JD55 (*MATa ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801 ubr1Δ:HIS3*) (34) and SGY3 (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1atel1-Δ2:LEU2*) (19a). Cells were grown in rich medium (YPD) or in synthetic media (SD) containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients, and 2% glucose. To induce the *P<sub>GAL</sub>* promoter, glucose was replaced by 2% galactose (SG medium). Transformation of *S. cerevisiae* was performed by the lithium acetate method (3).

The *ubr1Δ ate1Δ* double mutant AVY34 was constructed by replacing 93% of the *ATE1* open reading frame (ORF) (the first 470 codons) in strain JD55 (*ubr1Δ*) by the *LEU2* gene, through homologous recombination (42) with the introduced *LEU2* gene flanked on either side by 40 bp of *ATE1*-specific sequences. Mutants were selected on SD (lacking Leu and His) plates, and Leu<sup>+</sup> isolates were checked by PCR for the absence of *ATE1* and by colony assays (7, 8) for the absence of *Ate1p* and *Ubr1p* activity. High-copy-number pUB23-X plasmids expressing Ub-X-βgal proteins (see below) from *P<sub>GAL</sub>* in *S. cerevisiae* have been described elsewhere (4). Mouse *Ate1-1* and *Ate1-2* cDNAs (see below) were subcloned into the low-copy-number vector p414GAL1 (36), using the engineered *Bam*H1 (5') and *Xba*I (3') restriction sites, yielding plasmids pAT1 and pAT2. For localization assays with green fluorescent protein (GFP), cDNAs encoding mouse ATE1-1p or ATE1-2p were subcloned into the pEGFP-N1 N-terminal protein fusion vector (Clontech, Palo Alto, Calif.), using the engineered *Xba*I (5') and *Age*I (3') restriction sites, yielding plasmids pAT1-GFP and pAT2-GFP.

**Isolation of the mouse *Ate1-1* and *Ate1-2* cDNAs.** The 392-bp fragment of the mouse EST (expressed sequence tag) clone (accession no. AA415294), which was identified in GenBank through species walking (see Results), was used as a probe to screen a λgt10-based mouse cDNA library from MEL-C19 cells (Clontech), using standard procedures (3). Eight positive clones, whose inserts ranged from 0.5 to 1.6 kb, were analyzed by PCR and partial sequencing. The cDNA inserts of clones 3 and 8 were then subcloned into pBluescript II SK<sup>+</sup> (29) and sequenced on both strands. The resulting ORFs were identical except for a 129-bp internal region (see Results) (Fig. 2A). The deduced amino acid sequences of the mouse cDNA clones 3 and 8 were weakly but significantly similar to the deduced sequences of *Caenorhabditis elegans* and *S. cerevisiae* ATE1p (see Results) (Fig. 2B) and corresponded to nucleotides (nt) 699 to 1870 and 587 to 2099, respectively, in the subsequently produced full-length mouse *Ate1-1* and *Ate1-2* cDNAs (accession no. AF079096 and AF079097).

A human EST clone (accession no. AA503372) whose deduced amino acid sequence was highly similar to that of the partial mouse *Ate1* cDNA clone 8 was found in GenBank, using the partial mouse ATE1p sequence as a query. This EST clone was purchased from Genome Systems (St. Louis, Mo.) and sequenced



to obtain more of the 5'-proximal human *ATE1* sequence. Reverse transcription-PCR (RT-PCR) (3) was then carried out with poly(A)<sup>+</sup> RNA from mouse embryonic fibroblasts, using the mouse *Ate1*-specific reverse primer 5'-CCTTGGTAACAAACAGACTGGCTG-3' and the forward primer 5'-TCTCATAGACCGAGGATGGCGAAG-3', whose sequence was derived from the above human EST clone. The resulting PCR products, which appeared as a smear upon agarose gel electrophoresis, were ligated into the TA cloning vector (Invitrogen, San Diego, Calif.), and the ligation mixture was used as a template for PCR using a nested mouse *Ate1* primer 5'-CTGCAGCTGAGGCCCTGCTGCATCCG-3' and a vector-specific primer 5'-GTITCCCCAGTCACGAC-3'. This strategy yielded a single major DNA species (data not shown). We then applied 5'-RACE (rapid amplification of 5' cDNA ends) (3), using the above RT-PCR-derived sequence, to produce the full-length *Ate1* cDNA as previously described (19).

**Analysis of the mouse *Ate1* gene.** The mouse genomic DNA from L cells was used as a template for PCR, using the Expand high-fidelity PCR system (Boehringer, Indianapolis, Ind.) and exon-specific primers as previously described (29), to produce DNA fragments that together spanned ~4 kb of the mouse *Ate1* gene and contained the two alternative 129-bp *Ate1* exons. The regions encompassing exon/intron junctions were sequenced by using exon- and intron-specific primers. Thereafter, a strategy described earlier for the *Ubr1* gene (29) was used to screen, using a fragment of the mouse *Ate1* cDNA (nt 255-1139), a BAC (bacterial artificial chromosome)-based library of mouse genomic DNA fragments from strain 129svJ (Genome Systems), yielding one BAC clone containing the mouse *Ate1* gene.

**Isolation of the human, plant, and fly *ATE1* cDNAs.** Using the cloned mouse *ATE1p* sequences (see above) as queries, we identified in GenBank several significantly similar EST sequences from other organisms (data not shown). To determine whether these species also contained the two forms of *ATE1* mRNA, we isolated the corresponding *ATE1* cDNAs. RT-PCR (3) with poly(A)<sup>+</sup> RNA from human 293 cells and the primers 5'-CAATGGCATGTGGCACATTCC ATG-3' (specific for the human EST clone AA503372 [see above]) and 5'-CC ACAGGTACTGAATATGTATCCTG-3' (specific for the human EST clone AA195361) was carried out, yielding a 1.6-kb human *ATE1* cDNA fragment lacking the first 41 codons of the *ATE1* ORF. This fragment (a mixture of the two alternative cDNAs) was subcloned into the TA vector (Invitrogen) and sequenced on both strands. Full-length *Ate1* cDNAs from *Arabidopsis thaliana* and *Drosophila melanogaster* were isolated by RT-PCR as well, using total RNA from *A. thaliana* leaves, poly(A)<sup>+</sup> RNA from *D. melanogaster* embryos, and primers specific for the 5' and 3' ends of the corresponding ORFs. By using the strategy described above for the human *ATE1* cDNAs, the sequences of these primers were derived from the EST clones that were initially identified in GenBank through their similarity to the mouse *ATE1p* sequence, then purchased from Genome Systems, and sequenced prior to RT-PCR with the corresponding RNA preparations. The final human, plant, and fly *ATE1* cDNAs were sequenced on both strands.

**Assays of  $\beta$ -gal.** Colony assays for the *Escherichia coli*  $\beta$ -galactosidase ( $\beta$ gal) in *S. cerevisiae* were carried out by overlaying yeast colonies on SG plates with 0.5% agarose containing 0.1% sodium dodecyl sulfate (SDS), 4% dimethylformamide, and a 0.1-mg/ml solution of the chromogenic  $\beta$ gal substrate X-Gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside; Calbiochem, La Jolla, Calif.), followed by incubation for 1 to 2 h at 37°C. Quantitative assays for  $\beta$ gal in *S. cerevisiae* were carried out with whole-cell extracts, using another chromogenic  $\beta$ gal substrate, *o*-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG). Cells in a 5-ml culture ( $A_{600}$  of ~1) were pelleted by centrifugation and resuspended in 5 ml of buffer Z (60 mM Na<sub>2</sub>HPO<sub>4</sub>, 40 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM KCl, 1 mM MgSO<sub>4</sub>, 50 mM  $\beta$ -mercaptoethanol [pH 7.0]). After the  $A_{600}$  of the suspension was determined 50- or 100- $\mu$ l samples were diluted to 1 ml with buffer Z; 0.1% SDS (20  $\mu$ l) and CHCl<sub>3</sub> (50  $\mu$ l) were then added; the suspension was vortexed for 10 to 15 s and incubated for 15 min at 30°C, followed by the addition of 200  $\mu$ l of ONPG (4 mg/ml in buffer Z) and further incubation at 30°C, until a medium yellow color had developed, at which point the reaction was stopped by the addition of 1 M Na<sub>2</sub>CO<sub>3</sub> (0.4 ml). The mixture was centrifuged for 5 min at 1,100  $\times$  g, and the  $A_{420}$  and  $A_{500}$  of the samples were measured. The ONPG units

(U<sub>ONPG</sub>) of  $\beta$ gal activity were calculated as follows: U<sub>ONPG</sub> = 1,000  $\times$  [(A<sub>420</sub>) - (1.75  $\times$  A<sub>500</sub>)]/t  $\times$  v  $\times$  A<sub>600</sub>, where t and v were, respectively, the time of incubation (minutes) and the sample volume (milliliters) (3).

**Purification and N-terminal sequencing of X- $\beta$ gal proteins.** Extracts were prepared (using the liquid nitrogen procedure [3]) from *S. cerevisiae* AVY34 (*ubr1Δ ate1Δ*) cotransformed with a pUB23-X plasmid (4) (expressing Ub-X- $\beta$ gal) and either pAT1 (expressing mouse *Ate1-1p*) or pAT2 (expressing mouse *Ate1-2p*). Cultures were grown in SG to an  $A_{600}$  of ~1. Specific X- $\beta$ gal proteins (X = Asp or Cys) were purified by affinity chromatography on ProtoSorb lacZ (Promega, Madison, Wis.), a monoclonal anti- $\beta$ gal antibody coupled to agarose beads (Promega). X- $\beta$ gal proteins were further purified by electrophoresis on SDS-7% polyacrylamide gels and were electroblotted onto Immobilon-P<sup>SQ</sup> membranes (Millipore, Bedford, Mass.). N-terminal sequencing of 10 to 15 pmol of electroblotted X- $\beta$ gal was carried out for at least five cycles, using an Applied Biosystems 476A protein sequencer (Caltech Microchemistry Facility).

**Mouse cell cultures, transfection, and GFP localization.** NIH 3T3 cells (ATCC 1658-CRL) were grown as monolayers in Dulbecco's modified Eagle medium (GIBCO, Frederick, Md.) supplemented with 10% fetal bovine serum. Cells for GFP localization analyses were grown to ~15% confluence on glass coverslips for 24 h prior to transfection with either pAT1-GFP or pAT2-GFP, using Lipofectamine (GIBCO) and the manufacturer-supplied protocol. Cells were incubated for 5 h at 37°C in serum-free medium containing DNA and Lipofectamine. Thereafter an equal volume of medium containing 20% serum was added, and the cells were grown for another 12 to 20 h at 37°C. Cells were fixed with 2% formaldehyde in phosphate-buffered saline, and GFP fluorescence was visualized in a Zeiss Axiohot microscope.

**Northern hybridization.** Mouse multiple-tissue Northern blots containing 2  $\mu$ g of poly(A)<sup>+</sup> RNA per lane (Clontech) were probed with the <sup>32</sup>P-labeled 1.1-kb mouse *Ate1* cDNA (nt 638 to 1734), using the manufacturer-supplied protocol.

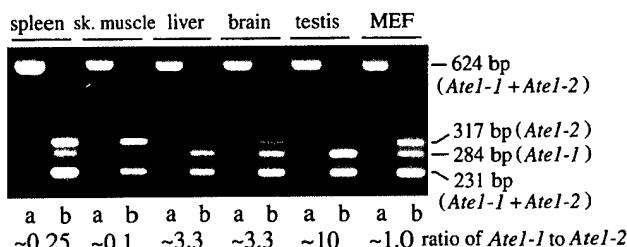
**Determination of the relative levels of *Ate1-1* and *Ate1-2* mRNAs.** Samples of total RNA isolated as described previously (3) from mouse spleen, skeletal muscle, liver, brain, testis, and embryonic fibroblasts were subjected to RT-PCR (28 cycles). The primers 5'-CAGTGGAGGATGCTGTTGACGGTGAC-3' and 5'-GTGCTCTGCCTCCAATGGTGAGCTG-3' were specific for the identical regions of *Ate1-1* and *Ate1-2* cDNAs that flanked the two 129-bp exons (see Results) which distinguished these cDNAs. The resulting 624-bp product (a mixture of the *Ate1-1* and *Ate1-2* cDNA fragments) was treated with *ScrFI*, which cuts at different sites within the two 129-bp exons, followed by a 2% agarose gel electrophoresis. This procedure made it possible to distinguish the *Ate1-1* and *Ate1-2* fragments. The ratios of the two forms of *Ate1* cDNA were determined by serial dilutions of the samples prior to gel electrophoresis.

**Nucleotide sequence accession numbers.** The nucleotide sequences reported in this paper were submitted to the GenBank/EMBL data bank and assigned accession no. AF079096 (mouse *Ate1-1* cDNA), AF079097 (mouse *Ate1-2* cDNA), AF079098 (human *Ate1-1* cDNA), AF079099 (human *ATE1-2* cDNA), AF079100 (*A. thaliana* *Ate1* cDNA), and AF079101 (*D. melanogaster* *Ate1* cDNA).

## RESULTS

**Identification of mouse *Ate1* cDNAs by species walking.** On the assumption that the sequences of R-transferases in different species might be sufficiently conserved to be detected by using the sequence of the only cloned R-transferase, *S. cerevisiae* *Ate1p* (7), we have been searching GenBank and related databases. No mammalian sequences in GenBank, including the EST sequences, had significant similarities to *S. cerevisiae* *Ate1p*. However, we did identify a nematode (*C. elegans*) ORF (accession no. Z21146) that exhibited similarity to yeast *Ate1p* (Fig. 2B) and then used the *C. elegans* sequence to identify a

**FIG. 2.** Two forms of the mouse *Ate1* cDNA and the ATE protein family. (A) The mouse *Ate1-1* and *Ate1-2* cDNAs and their products. The nucleotide sequences of *Ate1-2* identical to those of *Ate1-1* (everywhere except for the 129-bp region) are indicated by dashes. In the region of the alternative 129-bp exons of *Ate1-1* and *Ate1-2*, white-on-black and gray shadings highlight, respectively, identical and similar residues. The circled Cys residues are homologous to those that are important for the enzymatic activity of *S. cerevisiae* *Ate1p* (32). (B) The ATE protein family and the origins of the alternative 129-bp exons. Alignment of the sequences of mouse *ATE1-1p* (Mm-*Ate1-1*), *A. thaliana* *Ate1p* (At-*Ate1*), *D. melanogaster* *Ate1p* (Dm-*Ate1*), *C. elegans* *Ate1p* (Ce-*Ate1*), and *S. cerevisiae* *Ate1p* (Sc-*Ate1*) (accession no. J05404). Similar residues (gray) were grouped as follows: M, L, I, and V; D, E, N, and Q; R, K, and H; Y, F, and W; S, A, and T. The region encoded by the alternative 129-bp exons of mouse *Ate1* is highlighted by a thick line. Of the Cys residues that are conserved among all ATE proteins, the ones required and not required for the enzymatic activity of *S. cerevisiae* *Ate1p* (32) are indicated, respectively, by ▲ and ▽. The N-terminally truncated mouse *ATE1-1p* and *ATE1-2p* proteins which began at Met-42 (●) lacked the R-transferase activity (data not shown). The highly variable C-terminal regions of ATE proteins were omitted from the alignment. The sequences were aligned using PileUp program (Wisconsin Package; Genetics Computer Group, Madison, Wis.). Gaps (—) were introduced to optimize the alignment. The residue numbers are on the right of the sequences. The sequence of *C. elegans* *Ate1p* appears to lack the N-terminal region of other ATE proteins because of an error in defining the *Ate1* ORF in the genomic DNA sequence (accession no. Z21146). (C) Alignment of the 43-residue regions that are encoded by the alternative 129-bp exons in mammalian *Ate1*. Sequences shown: mouse (Mm-*Ate1-1* and Mm-*Ate1-2*), human (Hs-*Ate1-1* and Hs-*Ate1-2*), *D. melanogaster* (Dm-*Ate1*), *C. elegans* (Ce-*Ate1*), *A. thaliana* (At-*Ate1*), *S. pombe* (Sp-*Ate1*; accession no. Z99568), and *S. cerevisiae* (Sc-*Ate1*) (accession no. J05404). The degrees of identity and similarity of ATE proteins to the deduced amino acid sequences of the M8 (mouse *ATE1-1p*) or M3 (mouse *ATE1-2p*) exon of the mouse *Ate1* gene are indicated on the right. The residues conserved among all of the compared sequences are indicated above the alignment.



**FIG. 3.** Expression of *Atel-1* and *Atel-2* mRNAs in different mouse tissues and mouse embryonic fibroblasts (MEF). Two forms of *Atel* cDNA (Fig. 2A) were amplified in a single reaction by RT-PCR, using the same primers, to yield 624-bp fragments that included the region of the alternative 129-bp exons (see Materials and Methods). The 624-bp fragments were digested with *Sce*I, which produced a 231-bp fragment (a mixture of *Atel-1* and *Atel-2*), a 284-bp, *Atel-1*-specific fragment, and a 317-bp, *Atel-2*-specific fragment. The untreated (lanes a) and *Sce*I-treated (lanes b) samples from different mouse tissues were analyzed by electrophoresis in a 2% agarose gel. The ratio of the two forms of *Atel* mRNA, defined as the ratio of the 284-bp (*Atel-1*) fragment to the 317-bp (*Atel-2*) fragment, was determined by analyzing serially diluted samples and comparing the resulting band intensities (data not shown). sk., skeletal.

significantly similar EST sequence of *D. melanogaster* cDNA (accession no. AA391570). Finally, using the deduced amino acid sequence of the *Drosophila* EST clone as a probe, we identified a 467-bp mouse EST sequence (accession no. AA415294) that exhibited weak but significant similarity to the *Drosophila* sequence but no detectable similarity to *S. cerevisiae* *Ate1p*. On a chance that this 467-bp EST had been derived from the mouse *Ate1* cDNA, we used it to screen a mouse cDNA library and indeed isolated the putative mouse *Ate1* cDNAs (Fig. 2A).

**Alternative splicing results in two species of mouse *Ate1* cDNA containing distinct but homologous exons.** During the initial mouse cDNA library screening, we found that the cDNA clone 3 (nt 699 to 1870 of *Ate1-2* cDNA) was identical to the cDNA clone 8 (nt 587 to 2099 of *Ate1-1* cDNA), except for a 129-bp region whose deduced amino acid sequences were similar (31% identity; 61% similarity) (Fig. 2). The two full-length *Ate1* cDNAs (termed *Ate1-1* and *Ate1-2*), which were obtained by RT-PCR followed by 5'-RACE (see Materials and Methods), encoded proteins of identical length, 516 residues (59.2 kDa and pI of 8.14 versus 59.1 kDa and pI of 7.22) (Fig. 2A), that contained regions of similarity to the 57.8-kDa *Ate1p* of *S. cerevisiae* (Fig. 2B). RT-PCR (followed by subcloning) with RNAs from different mouse tissues also produced the two forms of *Ate1* cDNAs, indicating that the two species were in fact present in the initial RNA preparation (Fig. 3 and data not shown).

To determine whether both of the two 129-bp regions of the *Atel-1* and *Atel-2* cDNAs were a part of the *Atel* gene, and whether *Atel-1* and *Atel-2* were produced through alternative splicing, we analyzed the mouse *Atel* gene in the vicinity of its two 129-bp exons, using at first PCR and subsequently a BAC clone containing *Atel* (see Materials and Methods). The two 129-bp exons were located next to each other in the *Atel* gene (Fig. 4A). We also found that the 12-bp sequences around the splice acceptor sites of these exons (6 bp in the intron and 6 bp in the exon) were identical between the two exons (Fig. 4B), consistent with the alternative presence of these exons in the mature *Atel* mRNA. The exon-containing RT-PCR products from different mouse tissues appeared as a single major band retaining one of the two 129-bp exons (Fig. 3 and data not shown). Subcloning and analyses of these RT-PCR products yielded no other differentially spliced *Atel* cDNAs (for example, cDNAs retaining both or neither of the two 129-bp exons),

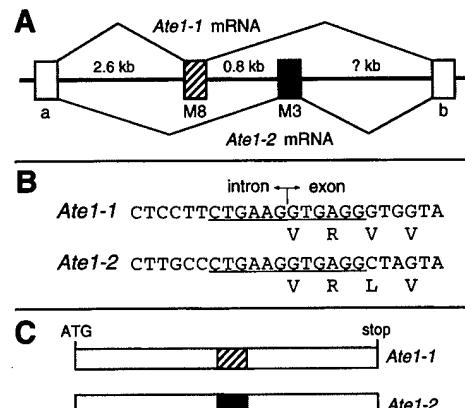


FIG. 4. The two forms of mouse *Ate1* mRNA are produced by alternative splicing. (A) The two alternative 129-bp exons are adjacent in the mouse *Ate1* gene. The thick line denotes genomic DNA; the striped and black rectangles denote the alternative 129-bp exons, M8 and M3 (see Materials and Methods); gray rectangles denote the flanking *Ate1* exons, of unknown sizes; thin lines denote the alternative splicing patterns that yield the two forms of *Ate1* mRNA. (B) The underlined 12 bp (6 bp in the intron and 6 bp in the exon) around the splice acceptor sites are identical between the two alternative 129-bp exons. (C) Scale diagrams of the two forms of mouse *Ate1* cDNAs. The alternative 129-bp exons M8 and M3 are indicated by the striped and black boxes, respectively.

suggesting that the splicing of *Ate1* pre-mRNA is tightly regulated to retain one and only one of the two alternative exons. Thus, the two forms of *Ate1* mRNAs are produced by a nearly unprecedented (see Discussion) splicing pathway which ultimately yields two proteins of identical size that bear two alternative, homologous but distinct 43-residue internal sequences (Fig. 4C).

#### The absence of alternative 129-bp exons from the plant and

**fly *Ate1* genes.** To explore the evolution of *Ate1*, and especially the phylogeny of its alternative 129-bp exons, we cloned the human, plant (*A. thaliana*), and fly (*D. melanogaster*) *ATE1* cDNAs (see Materials and Methods). Two forms of the human *ATE1* cDNA, termed *Hs-ATE1-1* and *Hs-ATE1-2*, were isolated from human 293 cells (the forms' molar ratio was about 1). However, only one form of the *Ate1* cDNA was isolated from either the leaves of *A. thaliana* (termed *At-Ate1*) or *D. melanogaster* embryos (termed *Dm-Ate1*), suggesting that the alternative 129-bp exons may not be present in the *Ate1* genes of plants and arthropods. The *A. thaliana* and *D. melanogaster* *Ate1p* proteins were, respectively, 629 and 477 residues long (71 and 55 kDa, with pIs of 6.0 and 8.4). Mouse *ATE1-1p* was 82, 38, and 42% identical (as well as 91, 57, and 61% similar) to human *ATE1-1p*, *A. thaliana* *Ate1p*, and *D. melanogaster* *Ate1p*, respectively (Fig. 2 and data not shown). *A. thaliana* *Ate1p* bore a 16-residue region containing exclusively Asp or Glu (data not shown).

We used RT-PCR and RNA preparations from *A. thaliana* and *D. melanogaster* to amplify the relevant regions of the corresponding *Atel* cDNAs. The resulting fragments were digested with restriction enzymes that recognize, in each species, exclusively the region that corresponds to the 129-bp exons of the mouse *Atel* cDNAs, and the products were analyzed by gel electrophoresis. The initial cDNA fragments of *A. thaliana* and *D. melanogaster Atel* completely disappeared after this treatment, in contrast to the homologous mouse cDNA fragment (which contained two distinct sequences of identical length), suggesting that the two alternative exons were absent from the *Atel* genes of plants and arthropods (Fig. 5A).

While this analysis was under way, complete sequences of

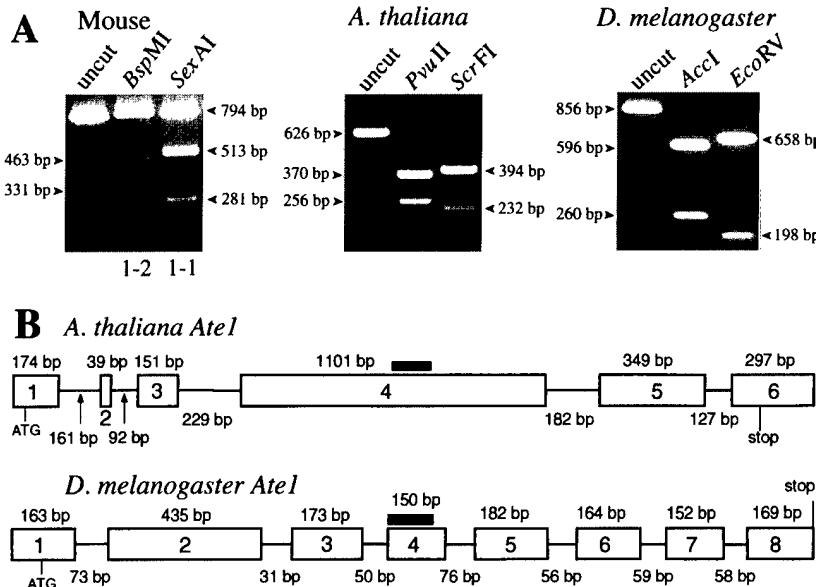


FIG. 5. Alternative splicing of *Ate1* pre-mRNA in mammals (the mouse) but in neither plants (*A. thaliana*) nor arthropods (*D. melanogaster*). (A) The relevant *Ate1* cDNA fragments from mouse (794 bp), *A. thaliana* (626 bp), and *D. melanogaster* (856 bp) were produced by RT-PCR (see Materials and Methods). The products were treated with the indicated restriction endonucleases that cut exclusively within the two alternative 129-bp exons of mouse *Ate1* cDNAs (*Bsp*MI for *Ate1*-2; *Sex*AI for *Ate1*-1) or within the corresponding regions of *A. thaliana* (*Pvu*II and *Scr*FI) and *D. melanogaster* (*Acc*I and *Eco*RV) *Ate1* cDNAs. (B) The *A. thaliana* and *D. melanogaster Ate1* genes. The exon-intron organization of these genes was deduced through comparisons of their cDNA sequences, determined in this work (see Materials and Methods), with the concurrently determined sequences of the corresponding genomic DNA regions (see text). The horizontal lines and rectangles denote, respectively, introns and exons, whose lengths are indicated below and above the line denoting introns. Thick horizontal lines indicate the regions of *A. thaliana* and *D. melanogaster* cDNAs that correspond to the alternative 129-bp exons of the mouse and human *Ate1* cDNAs (Fig. 2 and 4). The lengths of the *A. thaliana* and *D. melanogaster Ate1* genes are, respectively, ~3 and ~2.5 kb.

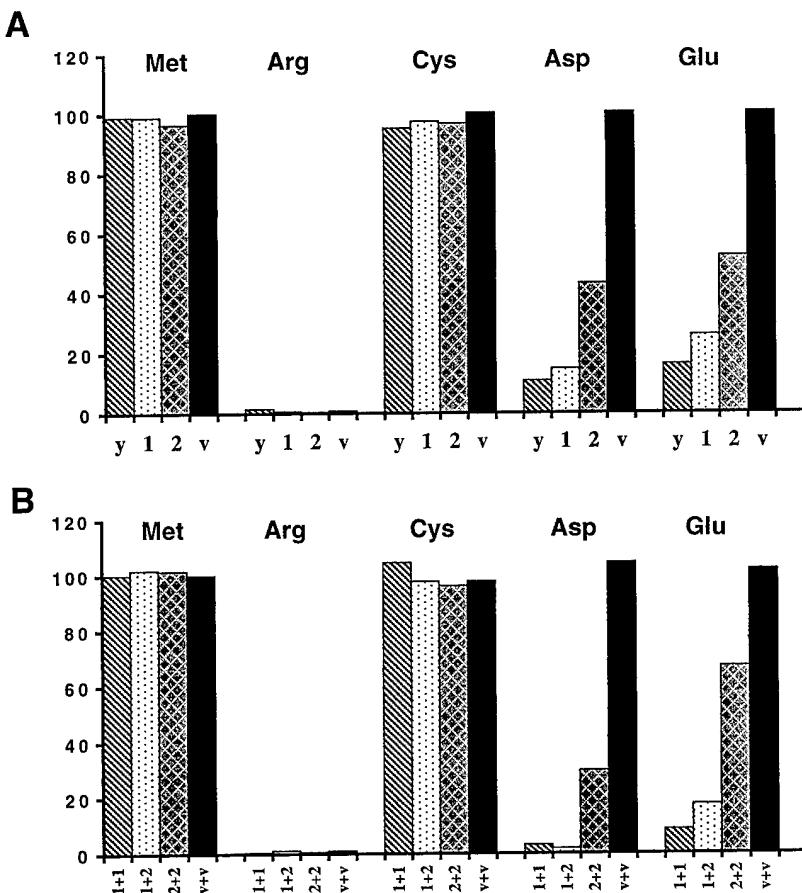
the *A. thaliana* and *D. melanogaster Ate1* loci, determined through the corresponding sequencing projects, were deposited in GenBank (accession no. AA005237 and accession no. AC004321, respectively). By comparing the cloned *Ate1* cDNAs (see Materials and Methods) and the corresponding genomic sequences of *A. thaliana* and *D. melanogaster*, we could deduce the organization of these *Ate1* genes. The results (Fig. 5B) directly confirmed the absence of the alternative homologous exons from *Ate1* of *A. thaliana* and *D. melanogaster*, in contrast to mammalian *Ate1*. The corresponding region of plant *Ate1p* is more similar to the exon-encoded sequence of mouse *ATE1-1p*, whereas in *Drosophila* this region is more similar to the alternative sequence of *ATE1-2p* (Fig. 2C). The corresponding regions of *S. cerevisiae* and *Schizosaccharomyces pombe* *Ate1p* are not preferentially similar to either of the two alternative exon-encoding sequences of mouse *ATE1p* (Fig. 2C).

**Mouse ATE1-1p and ATE1-2p can implement the Asp/Glu-specific subset of the N-end rule pathway but differ in activity.** To determine whether the two putative mouse R-transferases are in fact R-transferases and to compare their activities in an *in vivo* setting, we examined whether *ATE1-1p* and *ATE1-2p* could confer metabolic instability on Asp- $\beta$ gal and Glu- $\beta$ gal in *ate1* $\Delta$  *S. cerevisiae*. Asp and Glu are secondary destabilizing residues in the N-end rule (Fig. 1 and introduction). The test substrates Asp- $\beta$ gal and Glu- $\beta$ gal (produced through cotranslational deubiquitylation of Ub-Asp- $\beta$ gal and Ub-Glu- $\beta$ gal (4)) are short-lived in wild-type yeast (half-lives of ~3 and ~30 min, respectively) but long-lived (half-life of >20 h) in *ate1* $\Delta$  *S. cerevisiae* that lacks the *ATE1*-encoded yeast R-transferase (5, 7). Previous work (19, 33) has shown that the steady-state level of an X- $\beta$ gal protein is a sensitive measure of its metabolic stability.

*S. cerevisiae ate1* $\Delta$  cells were cotransformed with a pair of

plasmids that expressed one of the two putative mouse R-transferases, *ATE1-1p* or *ATE1-2p*, and one of several test substrates (as the corresponding Ub fusions): Asp- $\beta$ gal, Glu- $\beta$ gal, Arg- $\beta$ gal, Cys- $\beta$ gal, or Met- $\beta$ gal. Met and Cys are stabilizing residues in the yeast N-end rule; Arg is a primary destabilizing residue; Asp and Glu are secondary destabilizing residues (5, 56). Control tests included either the vector alone or a plasmid expressing *S. cerevisiae Ate1p*. The steady-state levels of X- $\beta$ gal proteins were determined by measuring the enzymatic activity of  $\beta$ gal in yeast extracts. Using this assay, we found that both forms of mouse *ATE1p* were able to confer metabolic instability on either Asp- $\beta$ gal or Glu- $\beta$ gal in *ate1* $\Delta$  *S. cerevisiae* (Fig. 6A). *ATE1-1p* and *ATE1-2p* destabilized Glu- $\beta$ gal much less than Asp- $\beta$ gal (Fig. 6A), consistent with Glu being a less destabilizing residue in the N-end rule than Asp, presumably because of less efficient arginylation of the N-terminal Glu by R-transferases (54). However, while the apparent destabilizing activity of the mouse *ATE1-1p* R-transferase was only slightly lower than that of *S. cerevisiae Ate1p* (expressed from the identical vector and promoter), the activity of mouse *ATE1-2p* was significantly lower than that of *ATE1-1p* (Fig. 6A).

We also asked whether the two forms of mouse *ATE1p* could influence each other's activity if they were coexpressed in the same cell (such an influence might be expected, for instance, if the active form of R-transferase were a dimer or if the two forms of R-transferase competed for binding to the same component of a pathway). *S. cerevisiae ate1* $\Delta$  cells were cotransformed with two plasmids bearing different selectable markers and expressing different combinations of *ATE1-1p* and *ATE1-2p* (1+1, 2+2, or 1+2), and also with a plasmid expressing one of the X- $\beta$ gal test proteins (X = Met, Arg, Cys, Asp, or Glu). Control cells were cotransformed with the two



**FIG. 6.** The two forms of mouse ATE1p can implement the Asp/Glu-specific subset of the N-end rule pathway. (A) Relative enzymatic activities of  $\beta$ -gal in *ate1* $\Delta$  *S. cerevisiae* transformed with plasmids expressing X- $\beta$ gal (as Ub-X- $\beta$ gal) test proteins (X = Met, Arg, Cys, Asp, or Glu) together with a plasmid expressing either yeast *ATE1* (denoted as y), mouse ATE1-1p (denoted as 1), mouse ATE1-2p (denoted as 2), or the vector alone (denoted as v). The N-terminal residues of X- $\beta$ gals in each set of experiments are indicated at the top. The activity of Met- $\beta$ gal in cells transformed with vector alone is taken as 100%. (B) The two forms of mouse ATE1p exhibit no cooperativity in mediating the degradation of X- $\beta$ gals. Shown are relative enzymatic activities of X- $\beta$ gals in *ate1* $\Delta$  *S. cerevisiae* strain cotransformed with plasmids expressing X- $\beta$ gals (Ub-X- $\beta$ gals) (X = Met, Arg, Cys, Asp, or Glu) and the combinations of plasmids expressing the following proteins: mouse ATE1-1p and ATE1-1p expressing X- $\beta$ gals (Ub-X- $\beta$ gals) (1+1), ATE1-1p and ATE1-2p (1+2), ATE1-2p and ATE1-2p (2+2), or two vector controls (v+v). One of the two vectors bore the *TRP1* marker and the other bore the *HIS3* marker (see Materials and Methods). Results are averages of four independent measurements, which differed by less than 10%.

vectors alone. The results (Fig. 6B) indicated that the total activities of the 1+1 and 1+2 combinations (measured as the extent of destabilization of Asp- $\beta$ gal or Glu- $\beta$ gal) were similar to each other and much higher than the total activities of 2+2 (Fig. 6B), consistent with the conjecture that the two forms of mouse R-transferase do not interact and that ATE1-1p is a more (possibly much more) active enzyme than ATE1-2p.

**Neither ATE1-1p nor ATE1-2p confers metabolic instability on Cys- $\beta$ gal.** Cysteine is a stabilizing residue in the yeast N-end rule but a secondary destabilizing residue in multicellular organisms such as mammals and amphibians (14, 18, 30, 54). The presence of two alternative regions in the two forms of mouse R-transferase (Fig. 4C) initially suggested that one R-transferase might be specific for N-terminal Asp and Glu, with the other specific for Cys. However, Cys- $\beta$ gal, which is long-lived in wild-type *S. cerevisiae* (5), remained long-lived in the presence of either ATE1-1p or ATE1-2p (Fig. 6A). This finding and, more directly, the results of amino acid sequencing (see below) suggest the existence of a mammalian tRNA-dependent enzyme (presumably a distinct R-transferase) (18) that mediates destabilizing activity of N-terminal Cys.

**Mouse ATE1-1p and ATE1-2p destabilize Asp- $\beta$ gal and Glu- $\beta$ gal through arginylation of their N-terminal residues.** To verify directly that mouse ATE1-1p and ATE1-2p in fact possess the R-transferase activity, we constructed the *ate1* $\Delta$  *ubr1* $\Delta$  *S. cerevisiae* double mutant AVY34, which lacked both R-transferase and N-recognin (E3), the main recognition component of the N-end rule pathway (see Materials and Methods). Consequently, N-terminal arginylation of a test protein in this mutant by an exogenous R-transferase would not result in degradation of the protein, thereby making it possible to isolate enough of the test protein for N-terminal sequencing. Strain AVY34 was transformed with pUB23-D (expressing Ub-Asp- $\beta$ gal) and also with either pAT1 (expressing ATE1-1p), pAT2 (expressing ATE1-2p), or vector alone and was grown in SG medium. Asp- $\beta$ gal proteins isolated from these transformants were subjected to N-terminal sequencing (see Materials and Methods). The results (Table 1) directly confirmed that both ATE1-1p and ATE1-2p possessed R-transferase activity. In agreement with the finding that ATE1-1p was more active than ATE1-2p in destabilizing Asp- $\beta$ gal in vivo (Fig. 6A), Asp- $\beta$ gal from cells expressing ATE1-1p was

TABLE 1. N-terminal sequencing of X- $\beta$ gal proteins isolated from *ate1* $\Delta$  *ubr1* $\Delta$  *S. cerevisiae* expressing different R-transferases

Substrate	Coexpressed protein	N-terminal sequence	Yield (%)
D-e <sup>K</sup> - $\beta$ gal	Vector alone	D-H-G-S-A-	
D-e <sup>K</sup> - $\beta$ gal	Mouse ATE1-1p	D-H-G-S-A-	~100
D-e <sup>K</sup> - $\beta$ gal	Mouse ATE1-2p	R-D-H-G-S-A-	~100
D-e <sup>K</sup> - $\beta$ gal	Mouse ATE1-2p	R-D-H-G-S-A-	~50
		D-H-G-S-A-	~50
C-e <sup>K</sup> - $\beta$ gal	Vector alone	C-H-G-S-A-	
C-e <sup>K</sup> - $\beta$ gal	Mouse ATE1-1p	C-H-G-S-A-	~20
C-e <sup>K</sup> - $\beta$ gal	Mouse ATE1-2p	C-H-G-S-A-	~20
		C-H-G-S-A-	~20

found to be completely arginylated, whereas Asp- $\beta$ gal from cells expressing ATE1-2p was arginylated to approximately 50% (Table 1).

We also determined, using the above procedure, whether mouse ATE1-1p or ATE1-2p could arginylate N-terminal Cys. Approximately 80% of Cys- $\beta$ gal isolated from *ate1* $\Delta$  *ubr1* $\Delta$  *S. cerevisiae* was found to be N-terminally blocked, presumably acetylated (Table 1). However, the rest of Cys- $\beta$ gal (~20%) bore the N-terminal sequence beginning with Cys and lacking N-terminal Arg, in agreement with the results of the in vivo Cys- $\beta$ gal degradation assays (Fig. 6 and Table 1). Thus, both ATE1-1p and ATE1-2p are apparently unable to utilize N-terminal Cys as a substrate in *S. cerevisiae*.

**ATE1-2p is exclusively cytosolic, whereas ATE1-1p is present in either the nucleus or the cytosol.** To determine the intracellular location of the two forms of mouse R-transferase,

we constructed fusions to the N terminus of GFP and transiently expressed them in NIH 3T3 cells. Whereas the free 26-kDa GFP was located in both the nucleus and the cytosol (data not shown), the 85-kDa ATE1-2p-GFP fusion was exclusively cytosolic in all of the many transfected cells examined (Fig. 7a to c). In contrast, the 85-kDa ATE1-1p-GFP (the alternative form of R-transferase that is much more active enzymatically than ATE1-2p) was found to be localized differently in different cells on the same coverslip, possibly depending on their cell cycle position and/or metabolic state. Specifically, in ~50% of the transfected cells, ATE1-1p-GFP was exclusively cytosolic (Fig. 7d and e), as was ATE1-2p-GFP (Fig. 7a to c), but in the other ~50% of cells, ATE1-1p-GFP was present in the nucleus as well and, moreover, appeared to be significantly enriched in the nucleus (Fig. 7f and g). Thus, the two 43-residue alternative regions in ATE1-1p and ATE1-2p confer overlapping but nonidentical intracellular distributions on the respective R-transferases.

While the nonuniformity of the ATE1-1p-GFP localization among mouse cells in a single culture remains to be understood, its preferential location in the nuclei of some cells is consistent with a high content of basic residues in its 43-residue region, in comparison to the alternative homologous region of ATE1-2p (Fig. 2C). (No sequences fitting the consensus sequences of known nuclear localization signals could be detected in the 43-residue region of ATE1-1p). In contrast to mouse R-transferases, *S. cerevisiae* Ate1p was shown to be located predominantly in the nuclei of yeast cells (56a).

**The ratio of *Ate1-1* to *Ate1-2* mRNA varies greatly among mouse tissues.** Northern hybridization, using the 1.1-kb mouse *Ate1* cDNA fragment (nt 638 to 1734) as a probe, detected a

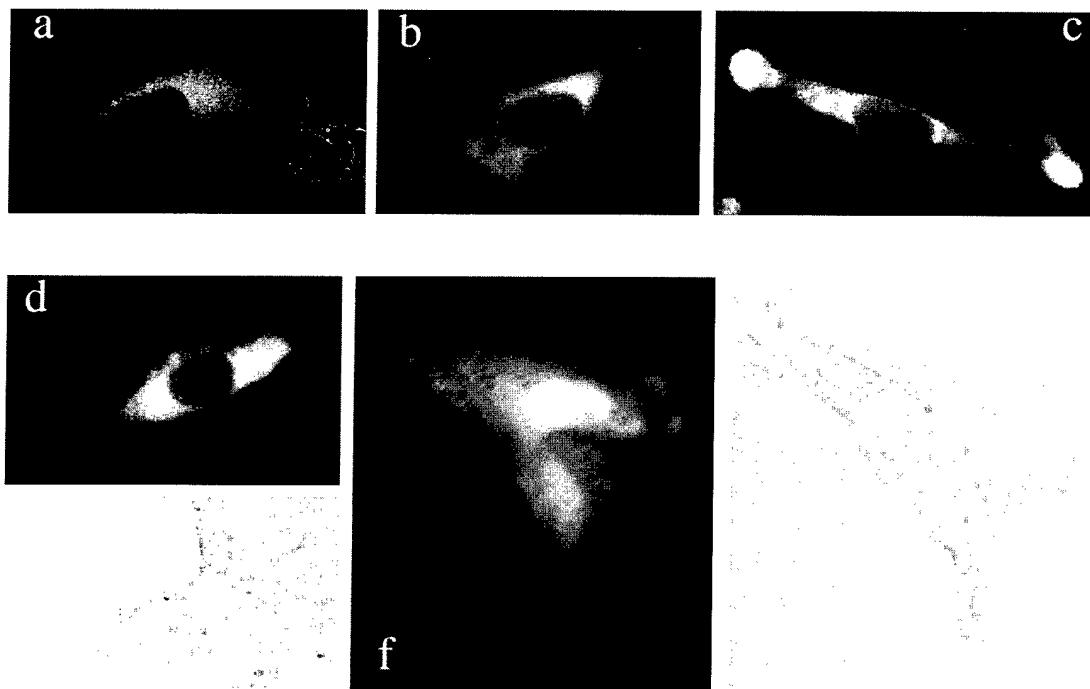


FIG. 7. Intracellular localization of mouse ATE1-1p and ATE1-2p. Shown are green (GFP) fluorescence (a to d and f) and phase-contrast (e and g) micrographs of mouse NIH 3T3 cells transiently transfected with ATE1-1p-GFP (d to g) or ATE1-2p-GFP (a to c) fusion proteins (see Materials and Methods). Panels a to c show different examples of the exclusively cytosolic localization of ATE1-2p. Regions around the nucleus and in the lamellar protrusions at the edges of a cell (c) exhibit higher GFP fluorescence, possibly because of a greater thickness of cells in these areas. Panels d plus e and f plus g show pairs of GFP fluorescence and phase-contrast pictures of cells that express ATE1-1p-GFP. The cell in panels d and e shows ATE1-1p-GFP in the cytosol but not in the nucleus. Cells in panels f and g contain ATE1-1p-GFP in both the cytosol and the nucleus, the latter being apparently enriched in ATE1-1p-GFP.

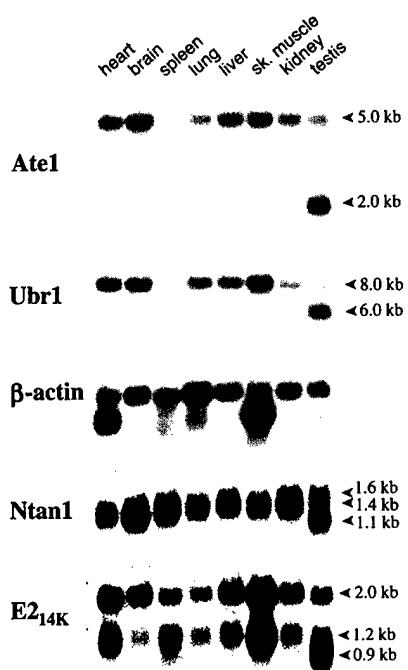


FIG. 8. Northern hybridization analyses of mouse *Atel* mRNA. The Northern blots of mRNA from different mouse tissues were probed with an *Atel* cDNA fragment (nt 638 to 1734) which can hybridize to both forms of *Atel* mRNA. A mouse β-actin cDNA probe was used for comparing the total RNA loads as described previously (19). The same blot was also hybridized with mouse *Ubr1* (29) (see the introduction). The apparent absence of *Atel* and *Ubr1* mRNAs from the spleen is an artifact of RNA degradation in this lane of the blot (data not shown). Also shown are the results of analogous Northern hybridizations of the mouse *Ntan1* cDNA, encoding the Asn-specific Nt-amidase (Fig. 1A), and *E2<sub>14K</sub>*, encoding the relevant Ub-conjugating (E2) enzyme (19). The approximate sizes of transcripts are indicated on the right.

single ~5.0-kb transcript (a mixture of the *Atel-1* and *Atel-2* mRNAs) in all of the mouse tissues examined except the testis, where the ~5-kb *Atel* mRNA was a minor one, the major species being ~2 kb (Fig. 8). Both ATE1p and the other targeting components of the mammalian N-end rule pathway are expressed ubiquitously (at various levels), and the testis-specific patterns of transcripts are characteristic for all of them as well (Fig. 8). The existence of the Y-chromosome-encoded, testis-specific variant of the Ub-activating (E1) enzyme (27, 35) suggests that the testis-specific modifications of the N-end rule pathway may be functionally relevant in spermatogenesis.

To determine the ratio of *Atel-1* to *Atel-2* mRNA in different mouse tissues or cells in culture, we employed RT-PCR, using sequence differences between the two alternative, homologous 129-bp exons to distinguish between them (see Materials and Methods) (Fig. 3). Approximately equal amounts of *Atel-1* and *Atel-2* mRNAs were present in mouse embryonic fibroblasts and in human 293 cells in culture (Fig. 3 and data not shown). However, the molar ratio of *Atel-1* to *Atel-2* mRNA was found to vary greatly among the mouse tissues: it was ~0.1 in the skeletal muscle, ~0.25 in the spleen, ~3.3 in the liver, and brain, and ~10 in the testis (Fig. 3). Thus, while the total expression of *Atel* (*Atel-1* plus *Atel-2*) varies by 2- to 4-fold among mouse tissues (Fig. 8), the difference in expression levels between *Atel-1* and *Atel-2* mRNAs can be as high as a 100-fold (the skeletal muscle versus the testis) (Fig. 3), suggesting that the two forms of R-transferase may be functionally distinct.

## DISCUSSION

The N-end rule pathway is one of several proteolytic pathways of the Ub system (23, 54, 55). Among the targets of the N-end rule pathway are proteins that bear destabilizing N-terminal residues. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by a specific amidase (6), to the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded R-transferase, to Arg, one of the primary destabilizing residues (7) (Fig. 1B). In mammals, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (18, 54) (Fig. 1).

In this work, we isolated cDNA encoding the mouse R-transferase, ATE1p, and found that this enzyme exists in two forms, termed ATE1-1p and ATE1-2p, which differ by containing one of the two alternative, homologous 43-residue regions. The two 516-residue R-transferases are produced from the mouse *Atel* gene by a pathway of alternative splicing that retains one or the other of the two homologous 129-bp exons. The presence of two adjacent, homologous, equal-length, and alternatively utilized exons in a gene (Fig. 4) is nearly unprecedented. To our knowledge, just one such case was described previously: the mouse κE2 enhancer-binding protein E12/E47 (37). The two κE2-binding proteins, E12 and E47, are produced through a switch between two alternative, equal-length exons, resulting in two helix-loop-helix DNA-binding proteins that differ in the ability to homodimerize. Specifically, E47 can bind to the κE2 enhancer either as a homodimer or as a heterodimer with MyoD, whereas E12 can bind as a heterodimer with MyoD but not as a homodimer (37).

We report the following major findings.

(i) Identification, through species walking, and isolation of the mouse cDNA encoding R-transferase (or ATE1p) have shown that mammalian ATE1p exists in two forms, ATE1-1p and ATE1-2p, which differ exclusively by one of the two alternative, homologous 43-residue regions (Fig. 2A).

(ii) The corresponding alternative 129-bp exons are adjacent in the mouse *Atel* gene. Moreover, the 12-bp sequences around the splice acceptor sites of these exons (6 bp in the intron and 6 bp in the exon) are identical between the two exons (Fig. 4). The splicing of *Atel* pre-mRNA proceeds in such way that one, and only one, of the alternative 129-bp exons is always retained in the mature *Atel* mRNA.

(iii) The human *ATE1* gene also contains the two alternative 129-bp exons, whereas the plant (*A. thaliana*) and fly (*D. melanogaster*) *Atel* genes encode a single form of ATE1p (Fig. 2 and 5). The corresponding 43-residue regions are significantly similar among all of the sequenced R-transferases, from *S. cerevisiae* to mammals (Fig. 2C). The set of *Atel* genes from mammals to yeast defines a distinct family of proteins, the ATE family. The splicing-derived alternative forms of R-transferase have evolved apparently after the divergence of the arthropod and vertebrate lineages.

(iv) Expression of the mouse *Atel-1* and *Atel-2* cDNAs in *atel1Δ S. cerevisiae*, and N-terminal sequencing of isolated X-βgal test proteins, was used to show that ATE1-1p and ATE1-2p could implement the Asp/Glu-specific subset of the N-end rule pathway and that they did so through the arginylation of N-terminal Asp or Glu in the test substrates (Fig. 6A and Table 1).

(v) While the destabilizing activity of the mouse ATE1-1p R-transferase is only slightly lower than that of *S. cerevisiae* R-transferase, the activity of mouse ATE1-2p is significantly

(possibly considerably) lower than that of ATE1-1p. This conclusion follows also from a comparison of the N-terminal arginylation of Asp- $\beta$ gal by the two R-transferases (Table 1). The results of coexpressing mouse ATE1-1p and ATE1-2p in the same *ate1* $\Delta$  yeast cells were consistent with the conjecture that R-transferase functions as a monomer (Fig. 6B).

(vi) Neither ATE1-1p nor ATE1-2p could confer instability on (or arginylate) Cys- $\beta$ gal in *ate1* $\Delta$  *S. cerevisiae* (Fig. 6A and Table 1). Cys is a stabilizing residue in yeast but a secondary destabilizing residue in the mammalian N-end rule (54). A distinct Cys-specific mammalian R-transferase suggested by these data remains to be identified.

(vii) Mouse ATE1-2p (tested as a GFP fusion) was exclusively cytosolic in mouse 3T3 cells, whereas ATE1-1p was localized differentially in different cells of the same (unsynchronized) culture: it was either exclusively cytosolic or present in both the cytosol and the nucleus (Fig. 7).

(viii) Mouse *Ate1* is a ubiquitously expressed gene. A single ~5-kb mRNA was present in all of the tissues examined except the testis, where the major *Ate1* transcript was ~2 kb in length (Fig. 8). The testis-specific differential expression patterns are also characteristic of the other targeting components of the mammalian N-end rule pathway, such as the *Ntan1*-encoded Asn-specific Nt-amidase and the *Ubr1*-encoded N-recognin (E3 $\alpha$ ) (19, 29).

(ix) The molar ratio of *Ate1*-1 to *Ate1*-2 mRNA varies up to a 100-fold among different mouse tissues (Fig. 3 and Results), suggesting a functional significance of the difference between the two R-transferases.

The region of ATE1p that corresponds to the two 129-bp mammalian *Ate1* exons has been significantly conserved throughout eukaryotic evolution, Tyr-296, Gln-297, and His-301 of the mouse ATE1-1p being among the most highly conserved residues (Fig. 2C). No putative members of the ATE family could be detected among the currently known prokaryotic ORFs. The most highly conserved region of R-transferases is an 82-residue stretch (residues 336 to 417) of mouse ATE1p; this region is 95, 76, and 63% identical to the corresponding regions of the human, *D. melanogaster*, and *A. thaliana* ATE1p, respectively (Fig. 2B). A Cys residue(s) is likely to be a component of the active site of R-transferase (31, 32). Among the five fully conserved Cys residues in proteins of the ATE family, four are located in the 56-residue N-terminal region (residues 23 to 78 of mouse ATE1p) (Fig. 2B). Conversion of some of these cysteines in *S. cerevisiae* Ate1p to alanines was found to decrease greatly the R-transferase activity of yeast Ate1p (16a, 32). Furthermore, derivatives of mouse ATE1-1p and ATE1-2p that lacked the first 42 residues were completely inactive in the yeast-based Asp- $\beta$ gal degradation assay of a kind described in Fig. 7 (data not shown). Finally, a 90-residue C-terminal truncation of *S. cerevisiae* Ate1p did not result in a major decrease of its R-transferase activity (16a). Thus, the active site of R-transferase is likely to encompass at least some of the above N-terminal cysteines.

Since the two mammalian R-transferases (Fig. 4C) are identical in size and, except for a 43-residue region, are identical otherwise as well, it is likely that the previously described (partially purified) mammalian R-transferases (12, 47) were in fact mixtures of Ate1-1p and Ate1-2p. On the other hand, fractionation of a crude R-transferase preparation from rabbit reticulocytes did yield, in addition to a major fraction of R-transferase, chromatographically distinct R-transferase fractions that were not investigated further (12). The ratio of Ate1-1p to Ate1-2p in rabbit reticulocytes is currently unknown.

A splicing-mediated switch that replaces the 129-bp *Ate1*

exon of ATE1-1p with the alternative 129-bp exon results in a protein, ATE1-2p, that has a significantly (possibly considerably) lower R-transferase activity (Fig. 6). In addition, ATE1-2p is unable to enter the nucleus (as a GFP fusion), in contrast to ATE1-1p (Fig. 7). Taken together with the finding that the expression ratio of the two *Ate1* mRNAs, *Ate1*-1 and *Ate1*-2, varies up to a 100-fold among different mouse tissues (Fig. 3 and Results), these data suggest that the two R-transferases are functionally distinct as well. Cited below are some of the possibilities that are consistent with the available evidence.

Mouse ATE1-2p has the R-transferase activity but arginylates, at steady state, only ~50% of Asp- $\beta$ gal in *ate1* $\Delta$  *S. cerevisiae*, in contrast to both ATE1-2p and *S. cerevisiae* Ate1p (Fig. 6 and Table 1). Moreover, the inefficient arginylation by ATE1-2p occurs in spite of its overexpression in *S. cerevisiae*. In contrast, the yeast Ate1p, which in wild-type *S. cerevisiae* is a weakly expressed protein (7), can quantitatively arginylate *in vivo* an overexpressed substrate such as Asp- $\beta$ gal (18). Thus, at a low level of expression (which is likely to be the case in the mouse), the ATE1-2p R-transferase may be, in effect, an inactive enzyme, in contrast to ATE1-1p. If so, ATE1-2p might act as an (indirect) inhibitor of the ATE1-1p function, for example, through a competition with ATE1-1p for the binding to a component of the targeting complex in the N-end rule pathway. (The apparent absence of such competition in *S. cerevisiae* [Fig. 6B] may result from the lack of binding by ATE1p to heterologous yeast proteins.) It is also possible that a large difference in activity between mouse ATE1-1p and ATE1-2p in yeast reflects not their different enzymatic activities in the mouse but a (physiologically irrelevant) differential recognition of an essential yeast cofactor such as Arg-tRNA. Direct comparisons of arginylation kinetics by the purified mouse and yeast R-transferases will be required to address this unlikely but unexcluded interpretation.

Another possibility is that ATE1-2p has a distinct enzymatic activity that has been missed by the current N-terminal arginylation assay (Fig. 6 and Table 1). For example, ATE1-2p might be able to arginylate an internal residue in a substrate protein. In vitro enzymological dissection of ATE1-1p and ATE1-2p will address this and related conjectures. Yet another possibility is that the alternative 43-residue regions of ATE1-1p and ATE1-2p confer different metabolic stabilities on the two R-transferases, the lower apparent activity of ATE1-2p in yeast being due at least in part to its shorter half-life. A test of this model in mouse cells requires antibodies specific for the alternative regions of the two R-transferases; preparation of such antibodies is under way.

In yeast, the N-end rule pathway is present in both the cytosol and the nucleus. The apparent exclusion of mouse ATE1-2p from the nucleus and the different ratios of *Ate1*-1 to *Ate1*-2 mRNA among the mouse tissues suggest that the rule book of the N-end rule pathway may be regulated differentially in the cytosol and the nucleus, through a cell-type-specific expression of the pathway's components that are located in one but not the other compartment.

Physiological substrates for either eukaryotic R-transferases (54) or their prokaryotic counterparts, L, F-transferases (1, 25, 45) are not known. The cloning and characterization of the first mammalian *Ate1* cDNAs and genes (Fig. 2), and the discovery of alternative splicing that yields mouse ATE1-1p and ATE1-2p (Fig. 4) should facilitate understanding of the functions of mammalian R-transferases, in part through the analysis of ATE1-1p and ATE1-2p enzymes and also because it is now possible to construct mouse strains that lack ATE1-1p and/or ATE1-2p.

## ACKNOWLEDGMENTS

The first two authors contributed equally to this work.

We thank Gary Hathaway of the Caltech Microchemistry Facility for the sequencing of X- $\beta$ gal proteins. We are grateful to Hai Rao, Glenn Turner, Fangyong Du, and Lawrence Peck for helpful suggestions and to Fangyong Du, Federico Navarro-Garcia, Hai Rao, and Youming Xie for comments on the manuscript.

This work was supported by grants DK39520 and GM31530 to A.V. from the National Institutes of Health.

## REFERENCES

1. Abramochkin, G., and T. E. Shrader. 1995. The leucyl/phenylalanyl-tRNA-protein transferase. Overexpression and characterization of substrate recognition, domain structure, and secondary structure. *J. Biol. Chem.* **270**:20621-20628.
2. Alagramam, K., F. Naidor, and J. M. Becker. 1995. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **15**:225-234.
3. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.). 1996. Current protocols in molecular biology. Wiley-Interscience, New York, N.Y.
4. Bachmair, A., D. Finley, and A. Varshavsky. 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**:179-186.
5. Bachmair, A., and A. Varshavsky. 1989. The degradation signal in a short-lived protein. *Cell* **56**:1019-1032.
6. Baker, R. T., and A. Varshavsky. 1995. Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. *J. Biol. Chem.* **270**:12065-12074.
7. Balzi, E., M. Choder, W. Chen, A. Varshavsky, and A. Goffeau. 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**:7464-7471.
8. Bartel, B., I. Wünnig, and A. Varshavsky. 1990. The recognition component of the N-end rule pathway. *EMBO J.* **9**:3179-3189.
9. Baumeister, W., J. Walz, F. Zühl, and E. Seemüller. 1998. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**:367-380.
10. Byrd, C., G. C. Turner, and A. Varshavsky. 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* **17**:269-277.
11. Chau, V., J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky. 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**:1576-1583.
12. Ciechanover, A., S. Ferber, D. Ganot, S. Elias, A. Hershko, and S. Arfin. 1988. Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. *J. Biol. Chem.* **263**:11155-11167.
13. Coux, O., K. Tanaka, and A. L. Goldberg. 1996. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* **65**:801-817.
14. Davydov, I. V., D. Patra, and A. Varshavsky. The N-end rule pathway in *Xenopus* oocyte extracts. *Arch. Biochem. Biophys.*, in press.
15. Dayal, V. K., G. Chakraborty, J. A. Sturman, and N. A. Ingoglia. 1990. The site of amino acid addition to posttranslationally modified proteins in regenerating rat sciatic nerves. *Biochim. Biophys. Acta* **1038**:172-177.
16. deGroot, R. J., T. Rümenapf, R. J. Kuhn, and J. H. Strauss. 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **88**:8967-8971.
- 16a. Du, F., and A. Varshavsky. Unpublished data.
17. Ferber, S., and A. Ciechanover. 1987. Role of arginine-tRNA in protein degradation by the ubiquitin pathway. *Nature* **326**:808-811.
18. Gonda, D. K., A. Bachmair, I. Wünnig, J. W. Tobias, W. S. Lane, and A. Varshavsky. 1989. Universality and structure of the N-end rule. *J. Biol. Chem.* **264**:16700-16712.
19. Grigoryev, S., A. E. Stewart, Y. T. Kwon, S. M. Arfin, R. A. Bradshaw, N. A. Jenkins, N. J. Copeland, and A. Varshavsky. 1996. A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. *J. Biol. Chem.* **271**:28521-28532.
- 19a. Grigoryev, S., and A. Varshavsky. Unpublished data.
20. Haas, A. J., and T. J. Siepmann. 1997. Pathways of ubiquitin conjugation. *FASEB J.* **11**:1257-1268.
21. Hershko, A. 1991. The ubiquitin pathway for protein degradation. *Trends Biochem. Sci.* **16**:265-268.
22. Hill, C. P., N. L. Johnston, and R. E. Cohen. 1993. Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl. Acad. Sci. USA* **90**:4136-4140.
23. Hochstrasser, M. 1996. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**:405-439.
24. Hondermarck, H., J. Sy, R. A. Bradshaw, and S. M. Arfin. 1992. Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. *Biochem. Biophys. Res. Commun.* **30**:280-288.
25. Ichetovkin, I. L., G. Abramochkin, and T. E. Shrader. 1997. Substrate recognition by the leucyl/phenylalanyl-tRNA protein transferase: conservation within the enzyme family and localization to the trypsin-resistant domain. *J. Biol. Chem.* **272**:33009-33014.
- 25a. Johnston, J., and A. Varshavsky. Unpublished data.
26. Kaiji, H., G. D. Novelli, and A. Kaiji. 1963. A soluble amino acid-incorporating system from rat liver. *Biochim. Biophys. Acta* **76**:474-479.
27. Kay, G. F., A. Ashworth, G. D. Penny, M. Dunlop, S. Swift, N. Brockdorff, and S. Rastan. 1991. A candidate spermatogenesis gene on the mouse Y chromosome is homologous to ubiquitin-activating enzyme E1. *Nature* **354**:486-489.
28. King, R. W., R. J. Deshaies, J. M. Peters, and M. W. Kirschner. 1996. How proteolysis drives the cell cycle. *Science* **274**:1652-1659.
- 28a. Kwon, Y. T., V. Denenberg, and A. Varshavsky. Unpublished data.
29. Kwon, Y. T., Y. Reiss, V. A. Fried, A. Hershko, J. K. Yoon, D. K. Gonda, P. Sangani, N. G. Copeland, N. A. Jenkins, and A. Varshavsky. 1998. The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **95**:7898-7903.
- 29a. Kwon, Y. T., and A. Varshavsky. Unpublished data.
30. Levy, F., N. Johnsson, T. Rümenapf, and A. Varshavsky. 1996. Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**:4907-4912.
31. Li, J., and C. Pickart. 1995. Inactivation of arginyl-tRNA protein transferase by a bifunctional arsenoxide: identification of residues proximal to arsenoxide site. *Biochemistry* **34**:139-147.
32. Li, J., and C. M. Pickart. 1995. Binding of phenylarsenoxide to Arg-tRNA-protein transferase is independent of vicinal thiols. *Biochemistry* **34**:15829-15837.
33. Madura, K., R. J. Dohmen, and A. Varshavsky. 1993. N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.* **268**:12046-12054.
34. Madura, K., and A. Varshavsky. 1994. Degradation of Gα by the N-end rule pathway. *Science* **265**:1454-1458.
35. Mitchell, M. J., D. R. Woods, P. K. Tucker, J. S. Opp, and C. E. Bishop. 1991. Homology of a candidate spermatogenesis gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1. *Nature* **354**:483-486.
36. Mumberg, G., R. Müller, and M. Funk. 1994. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**:5767-5768.
37. Murre, C., P. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**:777-783.
38. Ota, I. M., and A. Varshavsky. 1993. A yeast protein similar to bacterial two-component regulators. *Science* **262**:566-569.
39. Pickart, C. M. 1997. Targeting of substrates to the 26S proteasome. *FASEB J.* **11**:1055-1066.
- 39a. Rao, H., and A. Varshavsky. Unpublished data.
40. Rechsteiner, M., L. Hoffman, and W. Dubiel. 1993. The multicatalytic and 26S proteases. *J. Biol. Chem.* **268**:6065-6068.
41. Reiss, Y., and A. Hershko. 1990. Affinity purification of ubiquitin-protein ligase on immobilized protein substrates. *J. Biol. Chem.* **265**:3685-3690.
42. Rothstein, R. 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.* **194**:281-301.
43. Rubin, D. M., S. van Nocker, M. Glickman, O. Coux, I. Wefes, S. Sadis, H. Fu, A. Goldberg, R. Vierstra, and D. Finley. 1997. ATPase and ubiquitin-binding proteins of the yeast proteasome. *Mol. Biol. Rep.* **24**:17-26.
44. Scheffner, M., S. Smith, and S. Jentsch. 1998. The ubiquitin conjugation system, p. 65-98. In J.-M. Peters, J. R. Harris, and D. Finley (ed.), *Ubiquitin and the biology of the cell*. Plenum Press, New York, N.Y.
45. Shrader, T. E., J. W. Tobias, and A. Varshavsky. 1993. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl/phenylalanyl-tRNA-protein transferase gene *aat*. *J. Bacteriol.* **175**:4364-4374.
46. Sijst, A. J. A. M., I. Pilip, and E. G. Palmer. 1997. The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. *J. Biol. Chem.* **272**:19261-19268.
47. Soffer, R. L. 1980. Biochemistry and biology of aminoacyl-tRNA-protein transferases, p. 493-505. In D. Söll, J. Abelson, and P. R. Schimmel (ed.), *Transfer RNA: biological aspects*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
48. Solomon, V., V. Baracos, P. Sarraf, and A. Goldberg. When muscles atrophy, rates of ubiquitin conjugation increase, largely through activation of the N-end rule pathway. *Proc. Natl. Acad. Sci. USA*, in press.
49. Solomon, V., S. H. Lecker, and A. L. Goldberg. 1998. The N-end rule pathway mediates a major fraction of protein degradation in skeletal muscle. *J. Biol. Chem.* **273**:25216-2522.
50. Stewart, A. 1995. Trends in genetics nomenclature guide. Elsevier Science, Ltd., Cambridge, United Kingdom.
51. Stewart, A. E., S. M. Arfin, and R. A. Bradshaw. 1995. The sequence of porcine protein N-terminal asparagine amidohydrolase: a new component of the N-end rule pathway. *J. Biol. Chem.* **270**:25-28.
52. Taban, C. H., H. Hondermarck, R. A. Bradshaw, and B. Boilly. 1996. Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. *Experientia* **52**:865-870.
53. Tobias, J. W., T. E. Shrader, G. Rocap, and A. Varshavsky. 1991. The N-end rule in bacteria. *Science* **254**:1374-1377.

54. Varshavsky, A. 1997. The N-end rule pathway of protein degradation. *Genes Cells* 2:13-28.
55. Varshavsky, A. 1997. The ubiquitin system. *Trends Biochem. Sci.* 22:383-387.
56. Varshavsky, A., C. Byrd, I. V. Davydov, R. J. Dohmen, F. Du, M. Ghislain, M. Gonzalez, S. Grigoryev, E. S. Johnson, N. Johnsson, J. A. Johnston, Y. T. Kwon, F. Lévy, O. Lomovskaya, K. Madura, I. Ota, T. Rümenapf, T. E. Shrader, T. Suzuki, G. Turner, P. R. H. Waller, and A. Webster. 1998. The N-end rule pathway, p. 223-278. In J.-M. Peters, J. R. Harris, and D. Finley (ed.), *Ubiquitin and the biology of the cell*. Plenum Press, New York, N.Y.
- 56a. Wang, H. R., and A. Varshavsky. Unpublished data.
57. Wang, Y. M., and N. A. Ingoglia. 1997. N-terminal arginylation of sciatic nerve and brain proteins following injury. *Neurochem. Res.* 22:1453-1459.

# The E2–E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain

Youming Xie and Alexander Varshavsky<sup>1</sup>

Division of Biology, 147-75, California Institute of Technology,  
1200 East California Boulevard, Pasadena, CA 91125, USA

<sup>1</sup>Corresponding author  
e-mail:avarsh@its.caltech.edu

We dissected physical and functional interactions between the ubiquitin-conjugating (E2) enzyme Ubc2p and Ubr1p, the E3 component of the N-end rule pathway in *Saccharomyces cerevisiae*. The binding of the 20 kDa Ubc2p by the 225 kDa Ubr1p is shown to be mediated largely by the basic residue-rich (BRR) region of Ubr1p. However, mutations of the BRR domain that strongly decrease the interaction between Ubr1p and Ubc2p do not prevent the degradation of N-end rule substrates. In contrast, this degradation is completely dependent on the RING-H2 finger of Ubr1p adjacent to the BRR domain. Specifically, the first cysteine of RING-H2 is required for the ubiquitylation activity of the Ubr1p-Ubc2p complex, although this cysteine plays no detectable role in either the binding of N-end rule substrates by Ubr1p or the physical affinity between Ubr1p and Ubc2p. These results defined the topography of the Ubc2p-Ubr1p interaction and revealed the essential function of the RING-H2 finger, a domain that is present in many otherwise dissimilar E3 proteins of the ubiquitin system.

**Keywords:** E2/E3/N-end rule/proteasome/RING finger/ubiquitin

## Introduction

Ubiquitin (Ub) is a 76 residue protein whose covalent conjugation to other proteins, usually in the form of a multi-Ub chain, marks these proteins for processive degradation by the 26S proteasome, an ATP-dependent multisubunit protease (Hochstrasser, 1996; Varshavsky, 1997; Baumeister *et al.*, 1998; Hershko and Ciechanover, 1998; Scheffner *et al.*, 1998). Through either constitutive or conditional degradation of many intracellular proteins, the Ub-dependent proteolytic pathways regulate a multitude of biological processes, including the cell cycle, cell differentiation, apoptosis, DNA transcription, replication and repair, signal transduction, functions of the nervous system and stress responses, including the immune response (Hicke, 1997; Varshavsky, 1997; Peters *et al.*, 1998).

The conjugation of Ub to other proteins involves the formation of a thioester between the C-terminus of Ub and a specific cysteine of the Ub-activating (E1) enzyme. The Ub moiety of E1~Ub thioester is thereafter transesterified to a specific cysteine in one of several Ub-conjugating (E2) enzymes. The Ub moiety of E2~Ub thioester is

conjugated, via the isopeptide bond, to the ε-amino group of either a substrate's Lys residue or a Lys residue of another Ub moiety, the latter reaction resulting in a substrate-linked multi-Ub chain (Chau *et al.*, 1989; Pickart, 1997; Scheffner *et al.*, 1998). Most E2 enzymes function in complexes with proteins called E3s. The functions of E3s include the initial recognition of degradation signals (degrons) in the substrate proteins, with different E3s recognizing different classes of degrons. At least some E3s, specifically those containing the HECT domain, accept the Ub moiety from the associated E2~Ub thioester, forming an E3~Ub thioester and acting as a proximal donor of the Ub moiety to substrates selected by the E3 (Scheffner *et al.*, 1995; Nuber and Scheffner, 1999; Wang *et al.*, 1999).

One pathway of the Ub system is the N-end rule pathway (Bachmair *et al.*, 1986; Varshavsky, 1996). Among the targets of this pathway are proteins bearing destabilizing N-terminal residues. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues, in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase, to Arg, one of the primary destabilizing residues (Balzi *et al.*, 1990; Baker and Varshavsky, 1995). In mammals, the deamidation step is bifurcated, in that two distinct amidases specific, respectively, for N-terminal Asn or Gln, mediate the activity of tertiary destabilizing residues (Stewart *et al.*, 1995; Grigoryev *et al.*, 1996). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded E3 (N-recognin), the recognition component of the N-end rule pathway. In *S.cerevisiae*, Ubr1p is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues Phe, Leu, Trp, Tyr, Ile, Arg, Lys and His (Bartel *et al.*, 1990). Ubr1p contains at least three substrate-binding sites. The type 1 site is specific for basic N-terminal residues Arg, Lys and His. The type 2 site is specific for the hydrophobic residues Phe, Leu, Trp, Tyr and Ile (Reiss *et al.*, 1988; Gonda *et al.*, 1989; Baker and Varshavsky, 1991). Ubr1p also targets short-lived proteins such as Cup9p and Gpa1p, which lack destabilizing N-terminal residues, and are recognized by Ubr1p through its third substrate-binding site, the exact location and specificity of which remain to be determined (Madura and Varshavsky, 1994; Byrd *et al.*, 1998). Similar but distinct versions of the N-end rule pathway are present in all organisms examined, from bacteria to mammals. For a summary of the currently known functions of this pathway, see Kwon *et al.* (1999a,b).

The roles of E3 proteins in the E2, E3-dependent ubiquitylation of substrates are not understood in detail.

It has been shown that the *UBR1*-encoded N-recognin, the E3 of the N-end rule pathway, functions in a complex with Ubc2p (Rad6p) (Madura *et al.*, 1993; Watkins *et al.*, 1993), one of 13 E2 enzymes in *S.cerevisiae* (Hochstrasser, 1996). The polyacidic C-terminal tail of Ubc2p was found to be essential for the binding of Ubc2p to Ubr1p, but the Ubc2p-binding site of Ubr1p remained unknown. The mechanism of cooperation between E3 (Ubr1p) and E2 (Ubc2p) in forming a substrate-linked multi-Ub chain is unknown for either this or other Ub-dependent pathways.

We report two main results. (i) The Ubr1p–Ubc2p interaction is mediated by the basic residue-rich (BRR) region of Ubr1p and the polyacidic tail of Ubc2p. Interestingly, this high-affinity interaction is not essential for the activity of the N-end rule pathway. (ii) The RING-H2 domain of Ubr1p, adjacent to the BRR region, is strictly required for ubiquitylation of N-end rule substrates, even though this domain is not required for the binding of Ubr1p to either Ubc2p or N-end rule substrates. Because the RING-H2 and similar RING variants are a characteristic feature of many otherwise dissimilar E3 proteins, the ubiquitylation-enabling function of the Ubr1p RING-H2 identified in the present work is likely to be a general property of E3 components in the Ub system.

### Terminology

Ubiquitin whose C-terminal (Gly76) carboxyl group is covalently linked to another compound is called the *ubiquityl* moiety, the derivative terms being *ubiquitylation* and *ubiquitylated*. The term Ub refers to both free ubiquitin and the ubiquityl moiety. This nomenclature, which is also recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, brings Ub-related terms in line with the standard chemical terminology (Varshavsky, 1997).

## Results

### Dissection of the Ubr1p–Ubc2p interaction

To search for Ubc2p-interacting proteins, we employed the two-hybrid technique (Fields and Song, 1989; James *et al.*, 1996), using the 20 kDa Ubc2p as bait. A plasmid thus identified contained a fragment of the *S.cerevisiae* open reading frame (ORF) YLR024C, which was previously termed *UBR2* (Hochstrasser, 1996; Kwon *et al.*, 1998), because the predicted sequence of the 216 kDa Ubr2p was 22% identical and 46% similar to that of the 225 kDa, *UBR1*-encoded N-recognin, the E3 of the N-end rule pathway (Bartel *et al.*, 1990). Ubr2p is not essential for cell viability under standard growth conditions and does not target N-end rule substrates, but does compete with Ubr1p for the binding to Ubc2p (H.Rao, Y.Xie and A.Varshavsky, unpublished data).

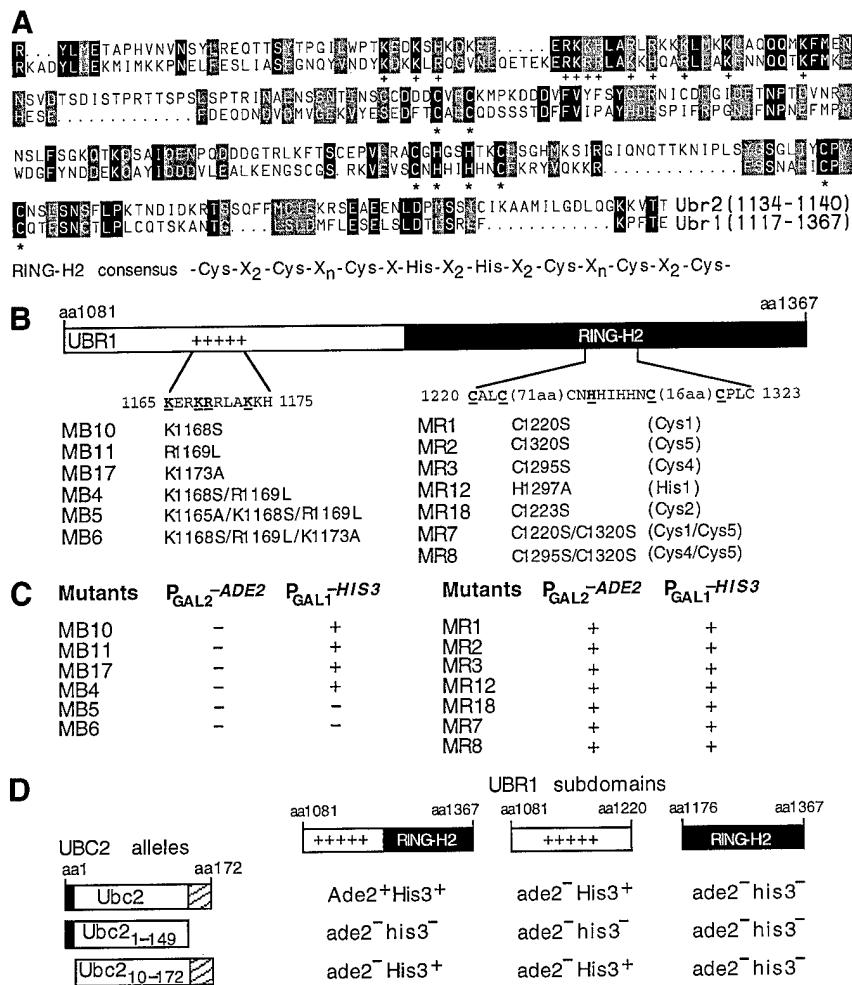
The cloned Ubc2p-interacting region of Ubr2p was a 287 residue fragment (positions 1134–1420). Sequence comparisons revealed two distinct motifs conserved between Ubr1p and Ubr2p in this region (Figure 1A). One was a basic residue-rich region, termed BRR, and the other a RING-H2 finger (Saurin *et al.*, 1996; Kwon *et al.*, 1998). The RING-H2 motif is a distinct variant of the previously defined Cys/His-rich RING motif, which is thought to be involved in protein–protein interactions (Borden and Freemont, 1996; Inouye *et al.*, 1997). Two-

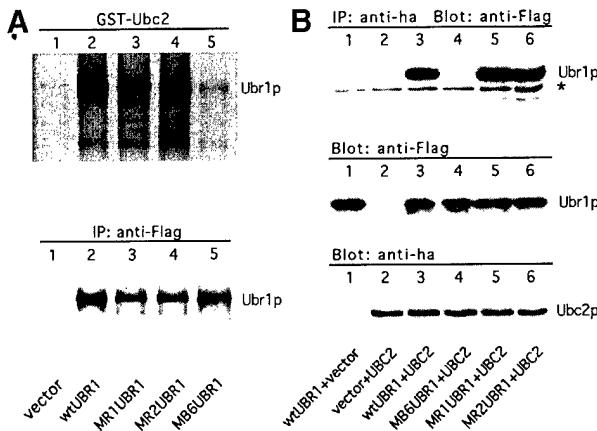
hybrid analyses indicated that Ubc2p interacted with the Ubr2p-homologous region of Ubr1p as well (positions 1081–1367) (Figure 1C and D). In addition, two-hybrid assays with a series of Ubr1p fragments fused to the Gal4p transcriptional activation domain demonstrated that the 287 residue region was the only detectable Ubc2p-binding site in Ubr1p (data not shown). We used site-directed mutagenesis to dissect the Ubc2p–Ubr1p interaction in this region of Ubr1p. Two sets of mutants were constructed (Figure 1B). The mutants of one set, termed MB (mutation in BRR), bore a missense mutation(s) in the BRR domain; the mutants of another set, termed MR (mutation in RING-H2) bore either Cys→Ser or His→Ala mutations in the RING-H2 finger. These derivatives of the 287 residue region of Ubr1p were tested for their binding to Ubc2p using a version of the two-hybrid assay that utilized two reporter genes, *ADE2* and *HIS3* (James *et al.*, 1996). This design increased the assay's affinity range, because a strong two-hybrid interaction was required to confer on the tester cells the Ade<sup>+</sup> phenotype, whereas a weak two-hybrid interaction sufficed to make the same tester cells His<sup>+</sup> (James *et al.*, 1996).

The replacement of a single basic residue at various positions in the BRR domain of Ubr1p with uncharged residues (e.g. the alleles MB10, MB11 and MB17) significantly weakened the interaction of Ubc2p with the 287 residue fragment of Ubr1p, as indicated by the Ade<sup>–</sup> His<sup>+</sup> phenotype of the tester strain, in comparison with the Ade<sup>+</sup> His<sup>+</sup> phenotype observed with the fragment's wild-type (wt) version (Figure 1C). When two basic residues of the BRR domain were converted to uncharged residues (MB4 allele), the resulting tester strain was still His<sup>+</sup>, but grew more slowly on histidine-lacking medium than did the tester strain bearing single-substitution Ubr1p mutants (Figure 1C; data not shown), suggesting a further weakening of the Ubc2p–Ubr1p interactions. With triple-substitution mutants in the BRR domain of Ubr1p (MB5 and MB6), the affinity of Ubc2p for the 287 residue region of Ubr1p decreased to levels undetectable with the two-hybrid assay (Figure 1C).

In contrast to these findings with the BRR domain, none of the four single Cys→Ser substitutions in the adjacent RING-H2 finger of Ubr1p impaired the interaction of Ubc2p with the 287 residue region of Ubr1p (Figure 1C, mutants MR1–MR3 and MR18). A His→Ala mutation in the RING-H2 finger was also without effect (Figure 1C, mutant MR12). Moreover, even the simultaneous Cys→Ser alterations at two positions of the RING-H2 finger failed to affect the Ade<sup>+</sup> His<sup>+</sup> phenotype of the tester strain (Figure 1C, mutants MR7 and MR8). The 287 residue region of Ubr1p was then subcloned into two fusions, which contained, respectively, the BRR domain alone (positions 1081–1220) and the RING-H2 finger alone (positions 1177–1367). The RING-H2 finger alone did not exhibit affinity to Ubc2p (Figure 1D). In contrast, the binding of Ubc2p to the BRR domain alone was readily detectable, but this interaction was weakened in comparison with the interaction between Ubc2p and the 287 residue region of Ubr1p containing both BRR and RING-H2 (Figure 1D).

Taken together, these results indicated that the positively charged BRR domain is the main region mediating the binding of Ubr1p to Ubc2p, through largely electrostatic





**Fig. 2.** The BRR domain of Ubr1p is essential for high-affinity binding to Ubc2p. (A) GST pull-down assays with wt Ubr1p and its mutants. FLAG-Ubr1p (denoted as wt UBR1) (lane 2), or FLAG-Ubr1p<sup>MB4</sup> (lane 3) or FLAG-Ubr1p<sup>MR1</sup> (lane 4), or FLAG-Ubr1p<sup>MR2</sup> (lane 5) were overexpressed in AVY301 (*ubr1Δ ubc2Δ*) cells (see Materials and methods). Cells were labeled with [<sup>35</sup>S]methionine for 30 min at 30°C, and the extracts were assayed using glutathione-agarose beads bound with GST-Ubc2p (top panel) or applied to immunoprecipitation with anti-FLAG monoclonal antibody (lower panel). (B) *In vivo* association of Ubc2p with wt Ubr1p, Ubr1p<sup>MB4</sup> and Ubr1p<sup>MR2</sup>, but not with Ubr1p<sup>MB6</sup>. Cell extracts from transformants expressing FLAG-Ubr1p alone (lane 1) or the N-terminally HA-tagged Ubc2p (HA-Ubc2p) alone (lane 2), or co-expressing HA-Ubc2p with either FLAG-Ubr1p (lane 3) or FLAG-Ubr1p<sup>MB4</sup> (lane 4), FLAG-Ubr1p<sup>MR1</sup> (lane 5) or FLAG-Ubr1p<sup>MR2</sup> (lane 6) were incubated with anti-HA antibody (see Materials and methods). Co-precipitated wt Ubr1p and its mutant derivatives were resolved by SDS-PAGE, followed by immunoblotting with anti-FLAG antibody (top panel). The asterisk marks a cross-reactive band. The expression levels of Ubr1p and Ubc2p in these transformants were examined by immunoblotting with, respectively, anti-FLAG (middle panel) and anti-HA antibodies (bottom panel). SDS-PAGE was carried out in 8% gels.

contiguous region within the BRR domain that is required for the binding of Ubc2p is likely to be significantly less than 140 residues.)

To determine whether basic residues of the BRR domain and the RING-H2 finger were essential for the binding of the full-length Ubr1p to Ubc2p, we first employed a pull-down assay with a fusion of Ubc2p to glutathione S-transferase (GST). Specifically, wt FLAG-Ubr1p (tagged at the N-terminus with the FLAG epitope), or FLAG-Ubr1p<sup>MB4</sup> (a triple-substitution mutant in the BRR domain; Figure 1B), or FLAG-Ubr1p<sup>MR1</sup> or FLAG-Ubr1p<sup>MR2</sup> (a single Cys→Ser substitution at Cys1 and Cys5, respectively; Figure 1B) was overexpressed in a *ubr1Δ ubc2Δ* *S.cerevisiae* strain and labeled with [<sup>35</sup>S]methionine. The N-terminal FLAG epitope did not impair the function of Ubr1p (A. Webster and A. Varshavsky, unpublished data). The extracts from labeled, Ubr1p-overexpressing cells (and the extract from cells carrying vector alone) were incubated with glutathione-agarose beads loaded with GST-Ubc2p, and the amounts of <sup>35</sup>S-labeled FLAG-Ubr1p and Ubr1p mutants recovered from the beads were determined by SDS-PAGE (Figure 2A, top panel). In parallel, the same extracts were immunoprecipitated with a monoclonal anti-FLAG antibody, to verify equality of the initial inputs of FLAG-Ubr1p and its mutants (Figure 2A, lower panel). In agreement with the results of the two-hybrid assays (Figure 1B and C), the binding

of FLAG-Ubr1p<sup>MB6</sup> to Ubc2p was barely detectable (Figure 2A, top panel, lane 5), whereas the binding of either FLAG-Ubr1p<sup>MR1</sup> or FLAG-Ubr1p<sup>MR2</sup> to Ubc2p was comparable to that of wt FLAG-Ubr1p (Figure 2A, top panel, lanes 2–4).

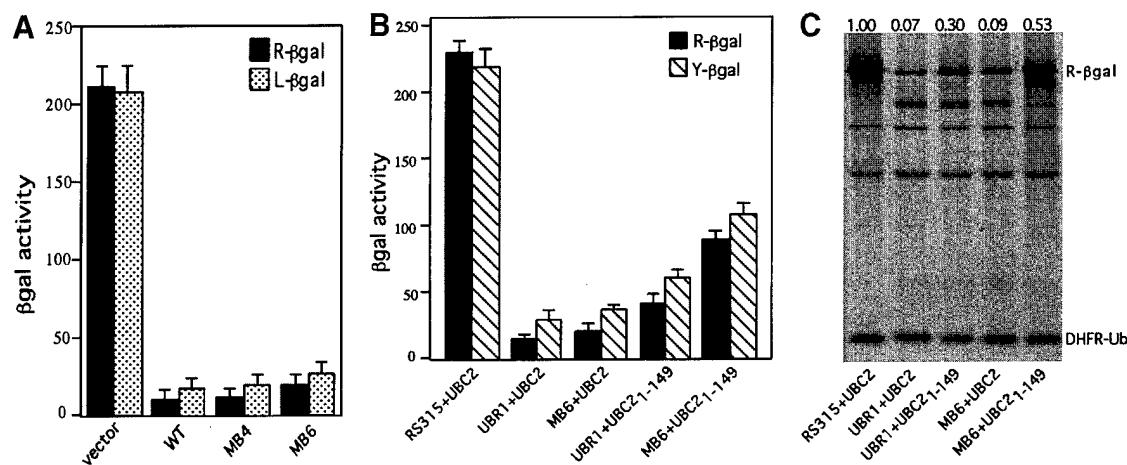
We also examined the *in vivo* association of wt Ubr1p and its mutants with Ubc2p, using co-immunoprecipitation-immunoblot assays. Either wt FLAG-Ubr1p, FLAG-Ubr1p<sup>MB6</sup>, FLAG-Ubr1p<sup>MR1</sup> or FLAG-Ubr1p<sup>MR2</sup> was co-overexpressed in a *ubr1Δ ubc2Δ* *S.cerevisiae* with Ubc2p bearing the N-terminal hemagglutinin (HA) epitope. Controls included congenic cells expressing FLAG-Ubr1p alone or Ubc2p alone. Proteins were immunoprecipitated from cell extracts with anti-HA antibody, followed by SDS-PAGE and immunoblotting with anti-FLAG antibody (Figure 2B, top panel). The expression levels of Ubr1p and Ubc2p in these transformants were monitored by immunoblotting the initial extracts with either anti-FLAG antibody (Figure 2B, middle panel) or anti-HA antibody (Figure 2B, bottom panel). In agreement with the results of two-hybrid and pull-down assays, the *in vivo* association of FLAG-Ubr1p<sup>MB6</sup> with Ubc2p was undetectable (lane 4), whereas both FLAG-Ubr1p<sup>MR1</sup> and FLAG-Ubr1p<sup>MR2</sup> bound Ubc2p as efficiently as wt FLAG-Ubr1p (lanes 3, 5 and 6). Thus, the BRR domain of Ubr1p is essential for the binding of Ubr1p to Ubc2p, whereas the adjacent RING-H2 finger of Ubr1p is not directly involved (Figures 1B, C and 2).

#### High affinity between Ubr1p and Ubc2p is not essential for degradation of N-end rule substrates

We asked whether the MB-type point mutations in the BRR domain of Ubr1p that virtually abolished the binding of Ubr1p to wt Ubc2p (Figures 1B, C and 2) affected the activity of the N-end rule pathway. wt Ubr1p and the BRR-domain mutants Ubr1p<sup>MB4</sup> and Ubr1p<sup>MB6</sup> (Figure 1C and D) were expressed in *ubr1Δ* cells from a low-copy plasmid and the native *P<sub>UBR1</sub>* promoter. The *ubr1Δ* cells also expressed an N-end rule substrate, either Arg-βgal or Leu-βgal, which are short-lived in *UBR1* cells (*t<sub>1/2</sub>* of ~2 and ~3 min, respectively) (Varshavsky, 1996). Arg-βgal and Leu-βgal were expressed as Ub-Arg-βgal and Ub-Leu-βgal, which were co-translationally deubiquitylated *in vivo* (Varshavsky, 1996). Previous work has also shown that the enzymatic activity of an N-end rule substrate such as X-βgal is a sensitive measure of the substrate's metabolic stability *in vivo* (Madura *et al.*, 1993; Kwon *et al.*, 1999a).

Both Arg-βgal and Leu-βgal were long-lived *in vivo* in the absence of Ubr1p (*t<sub>1/2</sub>* >20 h) ('vector' bars in Figure 3A; data not shown). As expected, Arg-βgal and Leu-βgal were degraded in the presence of wt Ubr1p (Figure 3A). Surprisingly, the steady-state levels of Arg-βgal and Leu-βgal in the transformants with Ubr1p<sup>MB4</sup> or Ubr1p<sup>MB6</sup> were similar to (or only slightly higher than) the levels of these proteins in transformants expressing wt Ubr1p (Figure 3A), indicating that high-affinity binding of Ubr1p to Ubc2p is not essential for the activity of the N-end rule pathway. Control experiments with *ubc2Δ* cells confirmed that the Ubr1p<sup>MB6</sup>-dependent degradation of N-end rule substrates was Ubc2p-dependent as well (data not shown).

Previous work with the tailless Ubc2p<sub>1–149</sub> mutant,



**Fig. 3.** High-affinity interaction between Ubr1p and Ubc2p is not essential for the activity of the N-end rule pathway. (A) The activity of Ubr1p is weakly affected by MB mutations. All *UBR1* alleles were expressed from the *P<sub>UBR1</sub>* promoter in a low-copy vector. Constructs expressing wt Ubr1p, Ubr1p<sup>MB4</sup> or Ubr1p<sup>MB6</sup>, or the empty vector were transformed into JD55 (*ubr1*<sup>-</sup>) cells expressing Arg- $\beta$ gal or Leu- $\beta$ gal. Measurements of  $\beta$ gal activity were used to assay the degradation of Arg- $\beta$ gal and Leu- $\beta$ gal. The values shown are the means of duplicate measurements of three independent transformants. Standard deviations are indicated above the bars. (B and C) High-affinity association of Ubr1p and Ubc2p is not essential for activity of the N-end rule pathway.  $\beta$ gal activity assays (B) and immunoprecipitations with anti- $\beta$ gal antibody (C) were used to compare the degradation of model substrates, either Arg- $\beta$ gal (B and C) or Tyr- $\beta$ gal (B). All *UBR1* and *UBC2* alleles were expressed from the native *P<sub>UBR1</sub>* and *P<sub>UBC2</sub>* promoters, respectively, in a low-copy vector. Different combinations of *UBR1* and *UBC2* alleles were transformed into *ubr1*<sup>-</sup> *ubc2*<sup>-</sup> cells expressing either Arg- $\beta$ gal, Tyr- $\beta$ gal (B) or DHFRha-Ub<sup>R48</sup>-Arg- $\beta$ gal (C) (see the text and Materials and methods). The measured ratio of Arg- $\beta$ gal to DHFR-Ub (the reference protein [(C), lane 1] was set at 1.00 and used to normalize the other ratios.

which did not bind to wt Ubr1p, led to the conclusion that Ubc2p<sub>1-149</sub> did not support the degradation of N-end rule substrates (Madura *et al.*, 1993), and implied that high-affinity binding of Ubc2p to Ubr1p was strictly essential for the pathway's function. However, another study reported a diminished but significant activity Ubc2p<sub>1-149</sub> (Watkins *et al.*, 1993). The latter conclusion was consistent with our findings about the BRR-region mutants of Ubr1p (Figure 3A). To address this discrepancy, we compared the activity of the N-end rule pathway in *ubr1*<sup>-</sup> *ubc2*<sup>-</sup> cells expressing different combinations of either wt Ubr1p or Ubr1p<sup>MB6</sup> with either wt Ubc2p or the tailless Ubc2p (Ubc2p<sub>1-149</sub>). *UBR1*, *UBC2* and their mutant alleles were expressed from their native promoters and a low-copy plasmid. The relative rates of degradation of Arg- $\beta$ gal and Tyr- $\beta$ gal ( $t_{1/2}$  of ~2 and ~10 min, respectively) (Bachmair *et al.*, 1986; Varshavsky, 1996) were assayed by measuring the activity of  $\beta$ gal (Figure 3B). With wt Ubc2p, the activity of the N-end rule pathway did not change significantly between the wt Ubr1p and Ubr1p<sup>MB6</sup> alleles (Figure 3B). Furthermore, in the presence of wt Ubr1p, the pathway's activity decreased significantly but not strongly upon the replacement of wt Ubc2p with its tailless derivative Ubc2p<sub>1-149</sub> (Figure 3B), in agreement with the finding of Watkins *et al.* (1993) but in contrast to the result of Madura *et al.* (1993). Interestingly, the activity of the N-end rule pathway was further decreased, but still not abolished, by combining Ubr1p<sup>MB6</sup> and Ubc2p<sub>1-149</sub> (Figure 3B).

To measure the relative metabolic stability of Arg- $\beta$ gal in a different way, we utilized a recently developed UPR (Ub/protein/reference) technique, which provides a reference protein (Lévy *et al.*, 1996, 1999). This method employs a linear fusion such as DHFRha-Ub<sup>R48</sup>-X- $\beta$ gal, in which Ub<sup>R48</sup>, containing Arg instead of Lys at position 48, is placed between a protein of interest (X- $\beta$ gal) and a reference protein moiety such as DHFRha (mouse

dihydrofolate reductase bearing the HA epitope tag) (Lévy *et al.*, 1996). DHFRha-Ub<sup>R48</sup>-X- $\beta$ gal is co-translationally cleaved by Ub-specific deubiquitylating enzymes (DUBs) (Wilkinson and Hochstrasser, 1998) after the last residue of Ub<sup>R48</sup>, producing equimolar amounts of X- $\beta$ gal and DHFRha-Ub<sup>R48</sup>. This way, the relative amounts of X- $\beta$ gal can be normalized against the (co-immunoprecipitated) DHFRha-Ub<sup>R48</sup> reference protein in the same sample. The UPR technique can therefore compensate for the scatter of immunoprecipitation yields, sample volumes, and other sources of sample-to-sample variation (Lévy *et al.*, 1996). *ubr1*<sup>-</sup> *ubc2*<sup>-</sup> cells expressing DHFRha-Ub<sup>R48</sup>-Arg- $\beta$ gal, and wt Ubr1p or Ubr1p<sup>MB6</sup> with wt Ubc2p or Ubc2p<sub>1-149</sub> were labeled with [<sup>35</sup>S]methionine for 30 min at 30°C. Cell extracts were immunoprecipitated with both anti- $\beta$ gal and anti-HA antibodies, followed by SDS-PAGE and the determination, using a PhosphorImager, of the ratio of Arg- $\beta$ gal (test protein) to DHFRha-Ub<sup>R48</sup> (reference protein) (Figure 3C). The ratio determined for *UBR1*-lacking cells was arbitrarily set at 1.00, and was used to normalize the same ratios from other transformants. The results (Figure 3C) were in agreement with those obtained by measuring the enzymatic activity of  $\beta$ gal (Figure 3A and B). Specifically, in the presence of wt Ubc2p and Ubr1p<sup>MB6</sup>, ~90% of Arg- $\beta$ gal was degraded under the assay's conditions, similar to the result with wt Ubc2p and wt Ubr1p. In cells expressing wt Ubr1p and Ubc2p<sub>1-149</sub>, ~70% of Arg- $\beta$ gal was degraded. With Ubr1p<sup>MB6</sup> and Ubc2p<sub>1-149</sub> the degradation of Arg- $\beta$ gal was decreased further, to ~50%, but still not abolished (Figure 3C).

Taken together, these results (Figure 3) indicated that high-affinity interactions between the E3 Ubr1p and the E2 Ubc2p are essential for normal levels of the pathway's activity. These data also showed that even much weaker interactions between mutant variants of Ubr1p and Ubc2p, the interactions that could not be detected by either two-hybrid or pull-down assays (Figures 1 and 2), were still

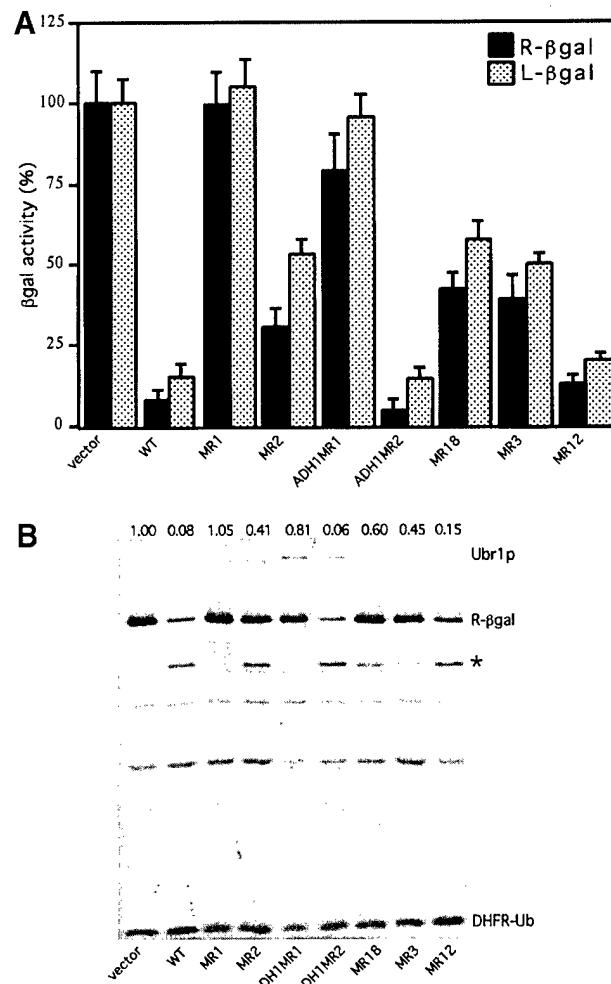
sufficient to sustain a large fraction of activity of the N-end rule pathway (Figure 3). We conclude that a transient interaction between Ubr1p and Ubc2p is functionally suboptimal but still sufficient for the multiubiquitylation of a substrate and for the substrate's delivery to the 26S proteasome.

#### The RING-H2 finger of Ubr1p is essential for degradation of N-end rule substrates

Given the proximity of the 104 residue RING-H2 finger to the BRR region of Ubr1p that mediates its binding to Ubc2p (Figure 1), we determined the effects of missense mutations in RING-H2 (Figure 1B) on the ability of Ubr1p to target N-end rule substrates for ubiquitylation and degradation. Since the structure of the folded RING-H2 domain is unknown, we constructed single missense mutants throughout this domain (Figure 1B). The sequences RING-H2<sup>MR1</sup>, RING-H2<sup>MR18</sup>, RING-H2<sup>MR3</sup> and RING-H2<sup>MR2</sup> contained single Cys→Ser substitutions at the Cys1, Cys2, Cys4 and Cys5 positions of RING-H2, respectively (Figure 1A and B). RING-H2<sup>MR12</sup> contained a His→Ala substitution at the His1 position of RING-H2 (Figure 1A and B). These RING-H2 variants were introduced into wt Ubr1. The resulting mutant proteins, Ubr1p<sup>MR1</sup>, Ubr1p<sup>MR2</sup>, Ubr1p<sup>MR3</sup>, Ubr1p<sup>MR12</sup> and Ubr1p<sup>MR18</sup>, were expressed in *ubr1Δ* cells from a low-copy plasmid, the native P<sub>UBR1</sub> promoter, and in the presence of either Arg-βgal or Leu-βgal, as described for the BRR-domain mutants of Ubr1p (Figure 3).

In striking contrast to the negligible effect of mutations in the RING-H2 finger of Ubr1p on the binding of Ubr1p to Ubc2p (Figures 1C, D and 2), all of the tested RING-H2 mutations except MR12 had a major inhibitory effect on the degradation of N-end rule substrates (Figure 4). Specifically, Ubr1p<sup>MR1</sup>, which bore a Cys→Ser substitution at Cys1 of the RING-H2 finger, was completely inactive as an E3 (N-recognition), in that the steady-state levels of Arg-βgal and Leu-βgal in cells expressing Ubr1p<sup>MR1</sup> were indistinguishable from the levels of these substrates in *ubr1Δ* cells, which lacked the N-end rule pathway (Figure 4). Other mutations of the RING-H2 finger impaired the N-end rule pathway strongly but incompletely, whereas the MR12 mutation (His→Ala at His1 of RING-H2) had a relatively weak effect (Figure 4).

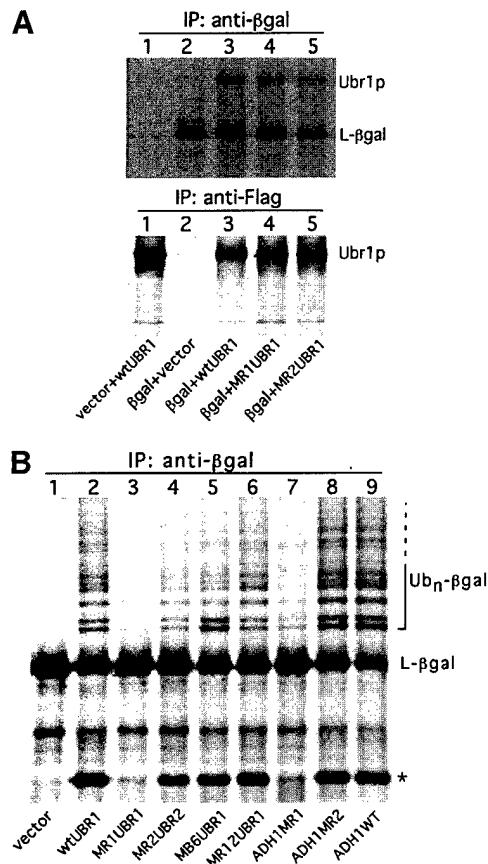
Previous work has shown that Ubr1p is a rate-limiting component of the N-end rule pathway, in that overexpression of Ubr1p accelerated the degradation of N-end rule substrates (Bartel *et al.*, 1990). We introduced the apparently inactive Ubr1p<sup>MR1</sup> and the partially active Ubr1p<sup>MR2</sup> mutations into wt Ubr1p bearing the N-terminal FLAG epitope and expressed the resulting proteins from the P<sub>ADH1</sub> promoter on a high-copy plasmid. Overexpression of the partially active Ubr1p<sup>MR2</sup> increased the activity of the N-end rule pathway in comparison with cells expressing Ubr1p<sup>MR2</sup> at wt levels (Figure 4). In contrast, the effect of overexpressing Ubr1p<sup>MR1</sup> was marginal, in that only ~4% of Leu-βgal and <20% of Arg-βgal were degraded, in comparison with virtually no degradation at physiological levels of Ubr1p<sup>MR1</sup> (Figure 4). Finally, the ~90 kDa *in vivo* cleavage product of βgal (denoted by an asterisk), which is characteristic of short-lived X-βgal substrates and is not produced in the absence of either



**Fig. 4.** Proteolysis by the N-end rule pathway requires the RING-H2 finger of Ubr1p. Methods described in the legend to Figure 3 and Materials and methods were used to assess the degradation of test substrates Arg-βgal (A and B) and Leu-βgal (A) in the presence of different *UBR1* alleles (see Figure 1B). All *UBR1* alleles were expressed from the native P<sub>UBR1</sub> promoter in a low-copy vector. The Ubr1p<sup>MR1</sup> (ADH1MR1) and Ubr1p<sup>MR2</sup> (ADH1MR2) mutants were also expressed from the P<sub>ADH1</sub> promoter in a high-copy vector. Constructs expressing wt Ubr1p and its mutants were transformed into JD55 (*ubr1Δ*) cells expressing the Arg-βgal or Leu-βgal [(A), βgal activity assays], or DHFR-Ub–Ub<sup>R48</sup>–Arg-βgal [(B), immunoprecipitation assays]. The activity of βgal in the absence of *UBR1* (vector alone) was set at 100%. The values shown are the means of duplicate measurements of three independent transformants. Standard deviations are indicated above the bars. In (B), the asterisk denotes the 90 kDa βgal cleavage product characteristic of the short-lived X-βgals (Bachmair *et al.*, 1986). Ubr1p co-immunoprecipitated with anti-βgal antibody is also indicated. The measured ratio of Arg-βgal to DHFR-Ub in the absence of Ubr1p [(B), lane vector] was set at 1.00 and used to normalize the other ratios.

Ubr1p or Ubc2p (Dohmen *et al.*, 1991), was nearly absent with Ubr1p<sup>MR1</sup> (Figures 4B and 5B); this result was consistent with negligible functional activity of Ubr1p<sup>MR1</sup>.

Previous work has shown that the N-end rule pathway controls the import of peptides through degradation of the homeodomain protein Cup9p, which inhibits the expression of the Ptr2p peptide transporter (Byrd *et al.*, 1998). We examined whether cells expressing Ubr1p<sup>MR1</sup> instead of wt Ubr1p were perturbed in their ability to import peptides by comparing the growth of *lys2Δ ubr1Δ* cells



**Fig. 5.** The RING-H2 finger of Ubr1p is required for ubiquitylation of N-end rule substrates. (A) The RING-H2 finger is not required for the binding of N-end rule substrates by Ubr1p. The UPR-based fusion DHFR-Ub<sup>R48</sup>-Leu- $\beta$ gal (see the text and Materials and methods) was co-expressed in AVY301 (*ubr1* $\Delta$  *ube2* $\Delta$ ) cells with the FLAG-tagged wt Ubr1p (lane 3), Ubr1p<sup>MR1</sup> (lane 4), Ubr1p<sup>MR2</sup> (lane 5) or vector alone (lane 2). The Ubr1p proteins were expressed from the PADH1 promoter in a high-copy vector. Lane 1, AVY301 cells expressing wt Ubr1p alone. Extracts from cells labeled with [<sup>35</sup>S]methionine were immunoprecipitated with anti- $\beta$ gal (top panel) or anti-FLAG antibodies (lower panel), followed by SDS-8% PAGE (see Materials and methods). (B) The RING-H2 finger of Ubr1p is required for ubiquitylation of N-end rule substrates. JD55 (*ubr1* $\Delta$ ) cells expressing Leu- $\beta$ gal and different amounts of Ubr1p or its mutants were labeled with [<sup>35</sup>S]methionine for 6 min at 30°C, followed by the addition of NEM to 50 mM and preparation of extracts by heating at 100°C in the presence of SDS (see Materials and methods). The extracts were precipitated with anti- $\beta$ gal antibody and fractionated by SDS-6% PAGE. Note the virtual absence of the multiubiquitylated Leu- $\beta$ gal species (denoted Ub<sub>n</sub>- $\beta$ gal on the right) from cells expressing Ubr1p<sup>MR1</sup> (lane 3). The asterisk denotes the 90 kDa  $\beta$ gal cleavage product characteristic of short-lived X- $\beta$ gals (Bachmair *et al.*, 1986).

expressing either wt Ubr1p, Ubr1p<sup>MR1</sup> or Ubr1p<sup>MB6</sup> on plates containing Lys-Ala dipeptide as the sole source of lysine. Under these conditions, inactivation of the N-end rule pathway is lethal, owing to the inability of cells (auxotrophic for lysine) to import the Lys-Ala dipeptide (Byrd *et al.*, 1998). Cells expressing Ubr1p<sup>MR1</sup>, which is virtually inactive as N-recognin, failed to grow on Lys-Ala dipeptide plates, whereas cells expressing either wt Ubr1p or Ubr1p<sup>MB6</sup> survived (data not shown), consistent with the effects of the MR1 and MB6 mutations on the activity of the N-end rule pathway (Figures 3 and 4).

We conclude that at least Cys1 of the RING-H2 finger in Ubr1p is essential for the E3 function of Ubr1p in the

N-end rule pathway, even though this cysteine is not required for the binding of Ubr1p to Ubc2p. Yet another insight from the pulse-immunoprecipitation assays with Arg- $\beta$ gal was that the anti- $\beta$ gal antibody co-precipitated the partially active FLAG-Ubr1p<sup>MR2</sup> and the inactive FLAG-Ubr1p<sup>MR1</sup> from the extracts of cells that over-expressed these Ubr1p mutants (Figure 4B). Thus, the functionally inactivating Cys1 $\rightarrow$ Ser1 mutation in the RING-H2 finger of Ubr1p<sup>MR1</sup> may not have impaired the binding of substrates by Ubr1p<sup>MR1</sup>. This conjecture is confirmed below.

#### ***RING-H2 of Ubr1p is not required for the binding of N-end rule substrates by Ubr1p***

To compare directly the affinities of Ubr1p and its derivatives for an N-end rule substrate, FLAG-Ubr1p, FLAG-Ubr1p<sup>MR1</sup> or FLAG-Ubr1p<sup>MR2</sup> were overexpressed in a *ubr1* $\Delta$  *ube2* $\Delta$  strain, which also expressed Leu- $\beta$ gal (produced from Ub-Leu- $\beta$ gal). Owing to the absence of Ubc2p, Leu- $\beta$ gal was long-lived in this strain, irrespective of the presence of Ubr1p (Dohmen *et al.*, 1991). Controls included congenic cells expressing FLAG-Ubr1p alone or Leu- $\beta$ gal alone. Cells were labeled with [<sup>35</sup>S]methionine for 30 min; proteins were immunoprecipitated from extracts with anti- $\beta$ gal antibody, followed by SDS-PAGE. Anti- $\beta$ gal co-immunoprecipitated Leu- $\beta$ gal and comparable amounts of either wt FLAG-Ubr1p, FLAG-Ubr1p<sup>MR1</sup> or FLAG-Ubr1p<sup>MR2</sup> (Figure 5A, upper panel, lanes 3–5). This co-immunoprecipitation was specific for Leu- $\beta$ gal, because no FLAG-Ubr1p was brought down with anti- $\beta$ gal in the absence of Leu- $\beta$ gal (Figure 5A, upper panel, lane 1). The expression levels of FLAG-Ubr1p, FLAG-Ubr1p<sup>MR1</sup> and FLAG-Ubr1p<sup>MR2</sup> were nearly identical, as could be shown by immunoprecipitating these proteins with anti-FLAG antibody (Figure 5A, lower panel). In contrast to the results with anti- $\beta$ gal antibody (Figure 5A, upper panel), Leu- $\beta$ gal was not co-precipitated with the FLAG-tagged Ubr1p proteins by anti-FLAG antibody (Figure 5A, lower panel). The probable explanation is that under these conditions, only free (overproduced) FLAG-Ubr1p proteins could be immunoprecipitated by anti-FLAG, because in the Leu- $\beta$ gal/FLAG-Ubr1p complex Leu- $\beta$ gal is bound to the type 2 substrate-binding site of Ubr1p, located in the N-terminal region of Ubr1p (A.Webster, M.Ghislain and A.Varshavsky, unpublished data). As a result, the N-terminus of Ubr1p could be shielded from recognition by the bulky  $\beta$ gal moiety bound to Ubr1p.

Taken together, these results (Figures 1–5) indicated that the MR1 (Cys1 $\rightarrow$ Ser1) mutation in the RING-H2 finger of Ubr1p virtually eliminated the E3 function of Ubr1p, but did not impair its affinity for either N-end rule substrates or the Ubc2p Ub-conjugating enzyme.

#### ***The RING-H2 finger of Ubr1p is required for ubiquitylation of N-end rule substrates***

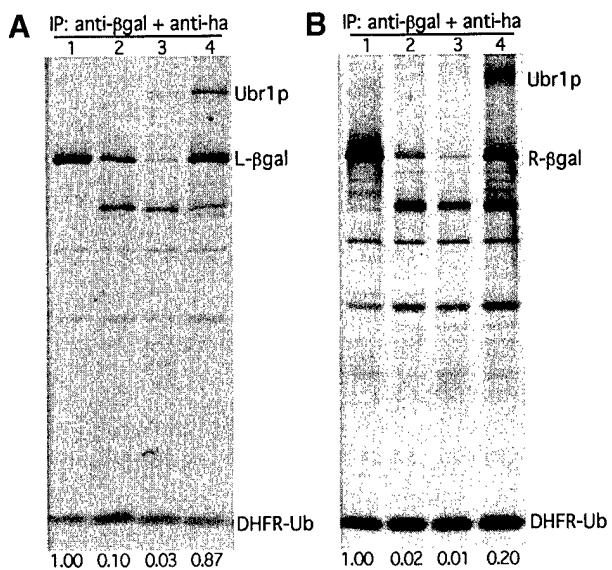
To examine ubiquitylation of an N-end rule substrate in cells containing mutant Ubr1p proteins, FLAG-Ubr1p<sup>MR1</sup>, FLAG-Ubr1p<sup>MR2</sup>, FLAG-Ubr1p<sup>MR12</sup>, FLAG-Ubr1p<sup>MB6</sup> (Figure 1B) and the wt FLAG-Ubr1p were expressed in a *ubr1* $\Delta$  *UBE2* strain, which also expressed Leu- $\beta$ gal (produced from Ub-Leu- $\beta$ gal). To minimize the extent of the degradation of Leu- $\beta$ gal during labeling, the duration

of the [ $^{35}$ S]methionine pulse was reduced to 6 min. At the end of the pulse, high concentrations of N-ethylmaleimide (NEM), the inhibitor of DUBs (Johnsson and Varshavsky, 1994), were employed to inhibit deubiquitylation after cell lysis.  $^{35}$ S-labeled, immunoprecipitated Leu- $\beta$ gal was fractionated by SDS-PAGE (Figure 5B).

Multiubiquitylated derivatives of Leu- $\beta$ gal were observed in the presence but not in the absence of wt Ubr1p, and the extent of ubiquitylation increased upon overexpression of FLAG-Ubr1p (Figure 5B, lanes 1, 2 and 9). However, virtually no ubiquitylated Leu- $\beta$ gal was observed with Ubr1p<sup>MR1</sup>, the degradation-inactive RING-H2 mutant of Ubr1p (Figure 5B, lane 3, compare with lanes 1 and 2). Cells overexpressing FLAG-Ubr1p<sup>MR1</sup> contained trace amounts of multiubiquitylated Leu- $\beta$ gal, in comparison with much higher levels of these species in cells overexpressing wt FLAG-Ubr1p (Figure 5B, lane 7, compare with lane 9). In agreement with the results of degradation tests (Figure 4), the partially degradation-defective RING-H2 mutant Ubr1p<sup>MR2</sup> yielded intermediate levels of the multiubiquitylated Leu- $\beta$ gal, and ubiquitylation was strongly enhanced upon overexpression of FLAG-Ubr1p<sup>MR2</sup> (Figure 5B, lanes 4 and 8). Ubr1p<sup>MR12</sup>, whose ability to mediate the degradation of N-end rule substrate was similar to that of wt Ubr1p (Figure 4), yielded the ubiquitylation pattern indistinguishable from that produced by wt Ubr1p (Figure 5B, lane 6). Consistent with the results of  $\beta$ gal activity assays (Figure 3), Ubr1p<sup>MB6</sup>, which bore three mutations in the BRR domain and was severely impaired in its binding to Ubc2p (Figures 1–3), yielded nearly wt levels of multiubiquitylated Leu- $\beta$ gal (Figure 5B, lane 5, compare with lane 2). Taken together, these results indicated that the RING-H2 finger of Ubr1p is specifically required for the Ubr1p, Ubc2p-dependent multiubiquitylation of N-end rule substrates, and that Cys1 of the RING-H2 finger is a functionally critical residue of this domain.

#### Dominant-negative effect of the ubiquitylation-inactive RING-H2 mutant of Ubr1p

The earlier gel filtration data suggested that Ubr1p functions as a monomer (Bartel, 1990). In addition, the immunoprecipitation of FLAG-Ubr1p with anti-FLAG antibody from extracts of cells that overexpressed both FLAG-Ubr1p and Ubr1p-HA did not co-immunoprecipitate Ubr1p-HA, also suggesting the monomeric configuration of Ubr1p (A. Webster and A. Varshavsky, unpublished data). We asked whether the functionally inactive Ubr1p<sup>MR1</sup> could exert a dominant-negative effect on the N-end rule pathway. Either wt Ubr1p or Ubr1p<sup>MR1</sup> (in addition to vector alone) was overexpressed from the P<sub>ADH1</sub> promoter on a high-copy plasmid in *UBR1* cells that also expressed the UPR-based fusions which produced Arg- $\beta$ gal (or Leu- $\beta$ gal) and equimolar amounts of DHFR-Ub<sup>R48</sup>, the reference protein. Cells were labeled with [ $^{35}$ S]methionine for 30 min, followed by immunoprecipitation of Leu- $\beta$ gal (Figure 6A) or Arg- $\beta$ gal (Figure 6B), as well as the reference protein DHFR-Ub<sup>R48</sup>. The ratios of X- $\beta$ gal to DHFR-Ub<sup>R48</sup> were measured, and thereafter normalized by the same ratios in congenic *ubr1* $\Delta$  cells (Figure 6A and B, lane 1). Whereas overexpression of wt Ubr1p in *UBR1* cells accelerated the degradation of Leu- $\beta$ gal and Arg- $\beta$ gal, overexpression of Ubr1p<sup>MR1</sup> strongly inhibited this degradation (Figure 6A and B, lanes 2 and 3, compare with lanes 4). The dominant-



**Fig. 6.** Dominant-negative effect of Ubr1p<sup>MR1</sup>, the RING-H2 mutant of Ubr1p. The P<sub>ADH1</sub> promoter-based, high-copy vector alone (lane 2), and its derivatives expressing the FLAG-tagged wt Ubr1p (lane 3) or Ubr1p<sup>MR1</sup> (lane 4) were transformed into JD52 (*UBR1*) cells expressing the UPR-based fusion DHFRha-Ub<sup>R48</sup>-Leu- $\beta$ gal (A) or DHFRha-Ub<sup>R48</sup>-Arg- $\beta$ gal (B). Cells were labeled with [ $^{35}$ S]methionine for 30 min at 30°C, followed by extraction, immunoprecipitation with a mixture of anti- $\beta$ gal and anti-HA antibodies, and SDS-12% PAGE, and quantitation. Lanes 1 in (A) and (B) show the results of the same tests with JD55 (*ubr1* $\Delta$ ) cells in the absence of any version of Ubr1p. The measured ratio of the test protein (Leu- $\beta$ gal or Arg- $\beta$ gal) to the reference protein (DHFRha-Ub<sup>R48</sup>) in lanes 1 was set at 1.00. The ratios determined in the other settings and normalized against the lane-1 ratios are shown below the lanes.

negative effect of overexpressed Ubr1p<sup>MR1</sup> was caused at least in part by the sequestration of endogenous Ubc2p, inasmuch as this effect could be partially reversed through overexpression of Ubc2p (data not shown). The sequestration of N-end rule substrates by Ubr1p<sup>MR1</sup>, which could bind to substrates indistinguishably from wt Ubr1p (Figure 5), was another likely (and independent) cause of the dominant-negative effect of Ubr1p<sup>MR1</sup> (Figure 6). The possibility that overexpression of Ubr1p<sup>MR1</sup> might also interfere with the downstream components of the Ub system, such as the 26S proteasome, was ruled out by the observation that the degradation of Ub-Pro- $\beta$ gal was not affected by the overexpression of Ubr1p<sup>MR1</sup> (data not shown). Ub-Pro- $\beta$ gal is a short-lived protein degraded by the UFD (Ub/fusion/degradation) pathway, which targets the non-removable (or slowly removable) N-terminal Ub moiety of a fusion protein (Johnson *et al.*, 1995; Koegl *et al.*, 1999).

## Discussion

We report two main results. First, we identified and dissected the E2 enzyme-binding site of Ubr1p, the E3 component of Ub-protein ligase in the N-end rule pathway of *S.cerevisiae*. This site, the BRR domain, is shown to bind the Ubc2p Ub-conjugating (E2) enzyme mainly through electrostatic interactions between the basic residues of BRR and the polyacidic C-terminal tail of Ubc2p. Although high-affinity interactions between the E3 Ubr1p and the E2 Ubc2p are essential for normal activity of the N-end rule pathway,

even much weaker interactions between mutant variants of Ubr1p and Ubc2p, the interactions which could not be detected by either two-hybrid or pull-down assays, were found to be sufficient for sustaining a large fraction of the pathway's activity. Secondly, we discovered the requirement for the RING-H2 finger of Ubr1p in the ubiquitylation of substrates mediated by the Ubr1p–Ubc2p complex. A single Cys→Ser missense mutation in the RING-H2 finger (which is adjacent to the BRR domain in Ubr1p) severely impairs ubiquitylation and degradation of N-end rule substrates, but does not impair the ability of Ubr1p to bind the substrates and Ubc2p. These findings defined the physical and functional aspects of the Ubc2p–Ubr1p interaction, and revealed a specific essential function of the RING-H2 finger. This domain is present not only in the members of the UBR family of E3s (Kwon *et al.*, 1998) but also in otherwise dissimilar E3 proteins of the Ub system.

#### **Physical and functional aspects of E2–E3 interaction**

Ubr1p, the 225 kDa E3 protein, was shown to interact with the 20 kDa Ubc2p E2 enzyme mainly, if not exclusively, through the BRR domain of <15 kDa. This interaction requires the polyacidic C-terminal tail of Ubc2p and has a strong electrostatic component, because conversions of single basic residues in the BRR domain into neutral ones resulted in comparable decreases in the affinity of Ubr1p for Ubc2p. Moreover, this affinity was decreased in rough proportion to the number of mutated basic residues in BRR (Figures 1 and 2). These results are consistent with the reported dependence of the Ubr1p–Ubc2p interaction on the length of the polyacidic tail of Ubc2p. Specifically, the truncated Ubc2p derivative that lacked eight of 23 residues of the polyacidic tail retained high affinity for Ubr1p, but further shortening of the Ubc2p polyacidic tail strongly decreased the affinity (Madura *et al.*, 1993).

Previous work suggested that the 9 residue N-terminal region of Ubc2p was essential for the binding of Ubc2p to Ubr1p (Watkins *et al.*, 1993). Our current data demonstrated that deletion of the first 9 residues of Ubc2p weakened but did not eliminate its interaction, in the two-hybrid assay, with the 287 residue fragment of Ubr1p that contained the BRR domain and the RING-H2 finger (Figure 1). It is not clear how the N-terminal 9 residues of Ubc2p contribute to the interaction with Ubr1p. Inasmuch as Arg6 and Arg8 of the folded Ubc2p are likely to interact through a hydrogen bond (Worthy lake *et al.*, 1998), it is possible that the effect of the 9 residue N-terminal deletion is indirect, stemming from a conformational destabilization of Ubc2p. Our finding of the 287 residue Ubr1p fragment that contains the BRR and RING-H2 domains and is sufficient for the high-affinity binding to Ubc2p (Figures 1 and 2) should facilitate the dissection of this E2–E3 interaction at the atomic level through co-crystallization of the previously solved Ubc2p (Worthy lake *et al.*, 1998) and the 287 residue fragment of Ubr1p.

#### **Function of the RING-H2 finger of Ubr1p**

One of the major results of this work is the discovery of an essential and specific requirement for the RING-H2 finger of Ubr1p in the ubiquitylation of N-end rule substrates mediated by the Ubr1p–Ubc2p complex. The RING-H2 motif is a distinct variant of the previously defined Cys/His-rich RING motif. In several proteins outside the Ub

system, RING-H2 has been shown to function as a site of protein–protein interactions (Borden and Freemont, 1996; Inouye *et al.*, 1997; Kamura *et al.*, 1999). In addition to the UBR family of E3 proteins (Kwon *et al.*, 1998), several other E3 components of the Ub system, including Apc11p, Roc1p (Rbx1p or Hrt1p), Roc2p and their homologs in multicellular eukaryotes, have also been found to contain RING-H2 or other variants of RING (Yu *et al.*, 1998; Zachariae *et al.*, 1998; Ohta *et al.*, 1999; Seol *et al.*, 1999; Skowyra *et al.*, 1999; Tan *et al.*, 1999).

The RING-H2 of Ubr1p is adjacent to the BRR domain, which mediates the physical interaction between Ubr1p and Ubc2p (Figures 1 and 2). A single Cys→Ser mutation at position 1 of the RING-H2 finger abolished ubiquitylation and degradation of N-end rule substrates, but did not impair the ability of the resulting mutant, Ubr1p<sup>MRI</sup>, to bind both the substrates and Ubc2p (Figures 4–6). The Cys→Ser mutations at other positions of the RING-H2 finger had similar but partial effects, while the His→Ala mutation at His1 position of RING-H2 had the smallest effect among the mutants tested (Figures 4–6). Thus, even though the Ubc2p E2 enzyme can interact with the E1 enzyme to yield the Ubc2p~Ub thioester, and can also produce isopeptide bond-mediated Ub–protein conjugates in the absence of E3 (Jentsch *et al.*, 1987), the synthesis of a multi-Ub chain by the Ubc2p–Ubr1p complex requires the presence of cysteine at position 1 of the RING-H2 finger in Ubr1p.

The Ub system of a cell contains a number of distinct E2 and E3 proteins (Hochstrasser, 1996; Varshavsky, 1997; Hershko and Ciechanover, 1998). An E3 protein of a Ub-dependent proteolytic pathway was initially presumed to function as a non-catalytic, substrate-recognition component of the corresponding Ub-protein ligase, which consists of an E3 and its associated E2 enzyme. Ubr1p (N-recognin) of the present work was the first E3 that has been shown to bind proteins bearing specific degradation signals (Bachmair *et al.*, 1986; Bartel *et al.*, 1990). More recently, it was found that in addition to recognizing specific substrates, at least some E3s, specifically those that contain the HECT domain (Scheffner *et al.*, 1993), also function as enzymes. An E3 of this class can accept the Ub moiety from an associated E2~Ub thioester, initially forming the E3~Ub thioester confined to a specific (and essential) cysteine of E3. The Ub moiety of this E3~Ub thioester is thereafter conjugated, through the isopeptide bond, either directly to a substrate protein or to the Ub moiety of a growing, substrate-linked multi-Ub chain (Scheffner *et al.*, 1995). Other classes of E3 proteins, including the SCF, APC and UBR families (Hershko and Ciechanover, 1998; Kwon *et al.*, 1998; Tyers and Willem, 1999), lack recognizable HECT domains, and have not thus far been demonstrated to form thioesters with Ub.

The mutations tested in the RING-H2 finger of Ubr1p (including the Cys1 mutation which abolished the ubiquitylation activity of the Ubr1p–Ubc2p complex) did not impair the ability of Ubr1p to bind the Ubc2p enzyme and N-end rule substrates (Figures 1–5). Thus, the specific functional effect of these mutations cannot be due to a global conformational destabilization of Ubr1p. We conclude that the Ubr1p (E3) component of the N-end rule pathway not only recognizes specific degradation signals in substrate proteins, but in addition directly

participates, through a mechanism that requires the RING-H2 domain, in the Ubc2p-mediated synthesis of a substrate-linked multi-Ub chain. In the first of possible models, a specific cysteine of RING-H2, perhaps the essential Cys1 residue (Figure 1), acts as an acceptor of the Ub moiety transesterified from the Ubc2p-Ub thioester, similar to the one previously described mechanism of the HECT-domain E3 called E6AP (Scheffner *et al.*, 1995; Huibregts *et al.*, 1998). Earlier experiments with an *in vitro* ubiquitylation system containing purified *S.cerevisiae* Ubr1p and other targeting components of the N-end rule pathway suggested the Ubc2p-dependent formation of a Ubr1p-Ub thioester (V.Chau and A.Varshavsky, unpublished data). Attempts to rigorously verify the functionally relevant presence of such a thioester in the completely defined, Ubr1p-based *in vitro* system have not been successful thus far (F.Du and A.Varshavsky, unpublished data). Another possible role of the cysteine(s) of the RING-H2 finger in Ubr1p may be to modify the immediate vicinity of the active cysteine (Cys88) of the Ubr1p-associated Ubc2p in a way that facilitates the multi-Ub synthesis by Ubc2p, in the absence of Ubr1p-Ub thioester formation. Physical proximity between the Ubc2p-binding BRR domain of Ubr1p and its RING-H2 domain (Figure 1A) is consistent with this mechanism. A third model posits the existence of an unidentified factor that binds to the RING-H2 finger of Ubr1p, and is required for the efficient synthesis of multi-Ub by the Ubr1p-Ubc2p complex. Jentsch and colleagues (Koegl *et al.*, 1999) have identified a factor, termed E4, which facilitates the synthesis of a multi-Ub chain linked to a substrate of the *S.cerevisiae* UFD pathway. This factor, encoded by the *UFD2* gene, is required for the *in vivo* degradation of Ub<sup>V76</sup>-Val-βgal, but is not required for the degradation of a much smaller and monomeric UFD substrate such as Ub<sup>V76</sup>-Val-DHFR; Ufd2p is also not required for the degradation of βgal-based N-end rule substrates (Johnson *et al.*, 1995). Whether the function of the RING-H2 finger of Ubr1p is mediated by a finger-binding protein analogous to E4 remains to be determined.

The RING-H2 finger and its variants are present in several otherwise dissimilar classes of E3 proteins in the Ub system. For example, the mammalian E3 encoded by the *Mdm2* gene and specific for the p53 substrate contains a putative RING finger (Elenbaas *et al.*, 1996); the conserved Cys464 residue of this domain was proposed to define an active site analogous to the Ub thioester-forming site of the HECT-domain E3 enzymes (Honda *et al.*, 1997). Roc1p (Rbx1p, Hrt1p), a recently identified subunit of the cell cycle-regulating Ub-protein ligase complex called SCF, contains a variant of RING-H2 in which Cys6 is replaced by Asp (Kamura *et al.*, 1999). Mutagenesis-based analyses suggested that the RING finger of Roc1p was required for the function of the SCF complex (Kamura *et al.*, 1999; Ohta *et al.*, 1999). The sequence of Roc1p (Rbx1p, Hrt1p) is highly similar to that of Apc11p, the RING-H2-containing subunit of another cell cycle-regulating Ub-protein ligase, called APC or cyclosome (Yu *et al.*, 1998; Zachariae *et al.*, 1998). The RING-H2 finger and its variants have also been found in other (possibly also E3) components of the Ub system, including Siah-1 and Der3p (Hrd1p), which are required, respectively, for the degradation of the DCC

protein (deleted in colorectal cancer) (Hu and Fearon, 1999) and of the misfolded luminal and integral membrane proteins (Bordallo *et al.*, 1998).

Further dissection of the RING domains in these and other E3 proteins will determine whether the ubiquitylation-enabling function of RING-H2 in Ubr1p of the N-end rule pathway is a general property of the E3 components in the Ub system.

## Materials and methods

### **Strains, media and $\beta$ -galactosidase assay**

The *S.cerevisiae* strains used were JD55 (MAT $\alpha$  *trp1*-Δ63 *ura3*-52 *his3*-Δ200 *leu2*-3, 112 *lys2*-801 *ubr1*Δ::HIS3); JD52 (MAT $\alpha$  *trp1*-Δ63 *ura3*-52 *his3*-Δ200 *leu2*-3, 112 *lys2*-801); AVY301 (a *ubc2*Δ::URA3 derivative of JD55); and PJ69-4A (a gift from P.James and E.Craig) (MAT $\alpha$  *trp1*-109 *leu2*-3, 112 *ura3*-52 *his3*-Δ200 *gal4*Δ *gal80*Δ *GAL2*-ADE2 *LYS2*::*GAL1*-HIS3 *met*::*GAL7*-lacZ). Cultures were grown in rich (YPD) or in synthetic media containing standard ingredients (Sherman, 1991) and either 2% glucose (SD medium), 2% galactose (SG medium) or 2% raffinose. Yeast were transformed as described previously (Chen *et al.*, 1992). For the induction of the  $P_{CUP1}$  promoter, CuSO<sub>4</sub> was added to a final concentration of 0.1 mM (Ghislain *et al.*, 1996). The enzymatic activity of βgal in yeast extracts, from cultures at OD<sub>600</sub> of 0.8–1.0, was determined as described previously (Baker and Varshavsky, 1991), using *o*-nitrophenyl-β-D-galactopyranoside (ONPG). The *Escherichia coli* strain used was DH5α.

### **Uses of the two-hybrid system**

The *UBC2* ORF was amplified by PCR using pKM1313 (Madura and Varshavsky, 1994) as the template DNA. The two primers were designed so that the PCR products contained a *Nco*I site immediately 5' to the start codon, and a *Sall* site immediately 3' to the stop codon. The *Nco*I-*Sall*-cut PCR fragment was subcloned into *Nco*I-*Sall*-cut PAS2 (Durfee *et al.*, 1993), yielding pAS2-UBC2, which expressed a fusion containing the DNA-binding domain of Gal4p upstream of the Ubc2p moiety. The construct was verified by nucleotide sequencing, and also by immunoblotting, in which the Gal4-Ubc2p protein was detected with an anti-Gal4 antibody (Santa Cruz Biotechnology, Santa Cruz, CA). Gal4-Ubc2p retained the ability of Ubc2p to rescue the UV-sensitivity defect of *ubc2Δ* cells. The PCR primers used were ggaaccatggagatgtccacacca-gatagaag and ctttgtcgactgatctgttcgtcgatcg. The two-hybrid DNA libraries and the host *S.cerevisiae* strain PJ69-4A were gifts from P.James and E.Craig (James *et al.*, 1996). PJ69-4A expressed three reporters from different inducible promoters: *P<sub>GAL1</sub>-HIS3*, *P<sub>GAL2</sub>-ADE2*, and *P<sub>GAL7</sub>-lacZ*. pAS2-UBC2 was transformed into PJ69-4A. Trp<sup>+</sup> transformants were then transformed with the two-hybrid libraries, with selection directly on SD (-trp, -leu, -ade) plates at 30°C. Ade<sup>+</sup> colonies were then tested on SD (-trp, -leu, -his) plates. Library plasmids were rescued from the Ade<sup>+</sup> His<sup>+</sup> transformants, and re-transformed into PJ69-4A expressing either pAS2-UBC2 or the control vector pAS2. The plasmids that induced *ADE2* and *HIS3* in the presence but not in the absence of pAS2-UBC2-encoded putative Ubc2p-interacting proteins, and were analyzed by partial sequencing. One clone thus identified encoded a 287 residue fragment (position 1134–1420) of the ORF YLR024C, which was previously termed *UBR2* (Hochstrasser, 1996; Kwon *et al.*, 1998), as described in Results.

ttccgtcgaccctcgtaaagggttgaaggcc. The PCR products were cut with *Bam*HI and *Sal*I, and subcloned into *Bam*HI-*Sal*I-cut pGADC1 (James *et al.*, 1996), yielding pGADC1UBR1<sub>1081-1367</sub>, which encoded a fusion of Ubr1p<sub>1081-1367</sub> to the transcriptional activation domain of Gal4p. *UBR1*<sub>1081-1220</sub>, encoding the Ubr1p fragment containing the BRR domain but not the RING-H2 finger, was constructed using the primers tacggatccatgagatttaaaacgtat and ttcgtcgacgcaggtaaaatctccgat, the above procedure, and yielding pGADC1UBR1<sub>1081-1220</sub>, which expressed a fusion of Ubr1p<sub>1081-1220</sub> to the transcriptional activation domain of Gal4. *UBR1*<sub>1177-1367</sub>, encoding the Ubr1p fragment containing the RING-H2 finger but not the BRR domain, was constructed using the primers aactggatcccaagggttctggcaag and ttcgtcgacgcaggtaaaatctccgattc, the above procedure, and yielding pGADC1UBR1<sub>1177-1367</sub>, which expressed a fusion of Ubr1p<sub>1177-1367</sub> to the transcriptional activation domain of Gal4. *UBR1* alleles containing point mutations in either the BRR domain or the RING-H2 finger were produced using the QuickChange mutagenesis kit (Stratagene, La Jolla, CA). Briefly, an ~850 bp *Eco*RI-*Bgl*II fragment from pGADC1UBR1<sub>1081-1367</sub> that encoded the BRR domain and the RING-H2 finger was subcloned into the *Eco*RI-*Bgl*II-cut LITMUS28 vector (New England Biolabs, Beverly, MA), yielding LITMUS28UBR1(*Eco*RI-*Bgl*II), which was subjected to site-directed mutagenesis. The resulting plasmids were cut with *Eco*RI and *Bgl*II, and the ~850 bp *Eco*RI-*Bgl*II fragments containing the intended point mutations were subcloned back into *Eco*RI-*Bgl*II-cut pGADC1UBR1<sub>1081-1367</sub>, yielding pGADC1UBR1MB10, pGADC1UBR1MB11, pGADC1UBR1MB17, pGADC1UBR1MB4, pGADC1UBR1-MB5, and pGADC1UBR1MB6; pGADC1UBR1MR1, pGADC1UBR1MR2, pGADC1UBR1MR3, pGADC1UBR1MR12, pGADC1UBR1-MR18, pGADC1UBR1MR7 and pGADC1UBR1MR8. All PCR products and inserts containing point mutations were verified by DNA sequencing on both strands. These mutants belonged to either the MB or the MR groups (Figure 1B).

#### UPR plasmids encoding N-end rule substrates

UPR plasmids were the yeast vector-based counterparts of the plasmids originally constructed for expression in mammalian cells (Lévy *et al.*, 1996). These plasmids, termed pR $\beta$ galUPR and pL $\beta$ galUPR, expressed, respectively, DHFR-Ub-Arg- $\beta$ gal and DHFR-Ub-Leu- $\beta$ gal from the P<sub>CUP1</sub> promoter. A *Bam*HI-*Sma*I insert of pBARUPR or pBALUPR was replaced with a 3.8 kb  $\beta$ gal-encoding insert from pUB23-M (Bachmair *et al.*, 1986), using *Bam*HI and *Scal*. pBARUPR and pBALUPR were derived from the low-copy, pRS314-based pBAM vector (Ghislain *et al.*, 1996). They expressed, respectively, DHFR-Ub-Arg-Ura3p and DHFR-Ub-Leu-Ura3p from the P<sub>CUP1</sub> promoter (A.Webster and A.Varshavsky, unpublished data), and were a gift from A.Webster.

#### Plasmids encoding Ubr1p, Ubc2p and their mutant derivatives

*UBR1* and its mutant alleles were expressed either from its natural promoter in a low-copy, pRS315-derived vector or from the P<sub>ADH1</sub> promoter in a high-copy, YEplac181-based plasmid. Briefly, pRS315 (Sikorski and Hietter, 1989) was cut with *Xba*I, blunted by Klenow Pol I, and thereafter self-ligated, yielding pRS315(-*Xba*I) lacking the *Xba*I site. pUBR1 (Bartel *et al.*, 1990) was cut with *Sal*I and *Pst*I. An ~8.1 kb insert containing the ORF and promoter of *UBR1* was subcloned into *Sal*I-*Pst*I-cut pRS315(-*Xba*I), yielding pRS315(-*Xba*I)UBR1. To produce mutant *ubr1* alleles, a ~2.9 kb *Bgl*II-*Xba*I fragment of *UBR1* from pRS315(-*Xba*I)UBR1 was subcloned into LITMUS28. The resulting LITMUS28UBR1(*Bgl*II-*Xba*I) was digested with *Hpa*I and *Sap*I, and an ~1.4 kb fragment was isolated. LITMUS28UBR1(*Bgl*II-*Xba*I) was also cut by *Sap*I, and an ~3.7 kb band was isolated. The 1.4 and 3.7 kb fragments were ligated with a set of ~600 bp *Sap*I-*Hpa*I fragments from the point mutation-containing pGADC1UBR1MB4, pGADC1UBR1MB6, pGADC1UBR1MR1, pGADC1UBR1MR2, pGADC1UBR1MR3, pGADC1UBR1MR12 or pGADC1UBR1MR18, yielding, respectively, pLITMUS28UBR1MB4, pLITMUS28UBR1-MB6, pLITMUS28UBR1MR1, pLITMUS28UBR1MR2, pLITMUS28-UBR1MR3, pLITMUS28UBR1MR12 and pLITMUS28UBR1MR18. A 2.9 kb *Bgl*II-*Xba*I fragment from each of these plasmids was subcloned into *Bgl*II-*Xba*I-cut pRS315(-*Xba*I)UBR1, yielding, respectively, pRS315(-*Xba*I)UBR1MB4, pRS315(-*Xba*I)UBR1MB6, pRS315(-*Xba*I)UBR1MR1, pRS315(-*Xba*I)UBR1MR2, pRS315(-*Xba*I)UBR1-MR3, pRS315(-*Xba*I)UBR1MR12 and pRS315(-*Xba*I)UBR1MR18. pNTFlagUBR1 expressed the full-length Ubr1p bearing the N-terminal FLAG epitope (MDYKDDDDK) from the P<sub>ADH1</sub> promoter in a high-copy, LEU2-containing vector. The MB6, MR1 or MR2 mutations (Figure 1B) were introduced into pNTFlagUBR1 by replacing its 2.9 kb

*Bgl*II-*Xba*I insert with its counterparts from pRS315(-*Xba*I)UBR1MB6, pRS315(-*Xba*I)UBR1MR1 or pRS315(-*Xba*I)UBR1MR2, yielding, respectively, pNTFlagUBR1MB6, pNTFlagUBR1MR1 and pNTFlagUBR1MR2.

*UBC2* and *UBC2*<sub>1-149</sub> alleles containing the native P<sub>UBC2</sub> promoter (~500 bp of the 5' region) were amplified by PCR. The *Sac*I-*Sal*I-cut PCR products were subcloned into *Sac*I-*Sal*I-cut p413GAL1 (Mumberg *et al.*, 1994). Both wt Ubc2p and Ubc2p<sub>1-149</sub> expressed from these vectors were able to complement the UV-sensitivity phenotype of *ubc2Δ* cells (data not shown). The PCR primers for the wt *UBC2* and the *ubc2*<sub>1-149</sub> were, respectively, acgtgagctcccataattgtttaagcg and ccttgtcactcgatctgttgcgtcg, and acgtgagctcccataattgtttaagcg and gctatgtcgactccaaggattcttacccgttc.

#### GST binding assays

The purified GST-Ubc2p fusion protein (~1  $\mu$ g) was diluted into 0.5 ml of the loading buffer (1% Triton X-100, 10% glycerol, 0.5 M NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 8.0), incubated with 10  $\mu$ l (bed volume) of the glutathione-agarose beads (Sigma, St Louis, MO) for 1 h at 4°C, and washed three times with 1 ml of the binding buffer (10% glycerol, 50 mM NaCl, 50 mM Na-HEPES pH 7.5). Washed beads were incubated with extracts from <sup>35</sup>S-labeled cells (~5  $\times$  10<sup>6</sup> CCl<sub>3</sub>COOH-insoluble counts/min/extract) containing the vector alone or a plasmid expressing one of the *UBR1* alleles, including wt FLAG-Ubr1p (specifically, pNTFlagUBR1, pNTFlagUBR1MB6, pNTFlagUBR1MR1 or pNTFlagUBR1MR2) at 4°C for 1 h. The incubation buffer contained 1 mg/ml of the ovalbumin carrier (Sigma, St Louis, MO) in the binding buffer. Beads were washed three times with the binding buffer. Bound proteins were eluted with the loading buffer containing 2% SDS and fractionated by SDS-8% PAGE.

#### Co-immunoprecipitation-immunoblotting

Ubc2p bearing the N-terminal HA epitope was expressed from the P<sub>GAL1</sub> promoter in a low-copy number vector p416GAL1 (Mumberg *et al.*, 1994). AVY301 cells (*ubr1Δ ubc2Δ*) containing a plasmid expressing HA-tagged Ubc2p or a plasmid expressing FLAG-Ubr1p, or plasmids expressing HA-tagged Ubc2p and one of the *UBR1* alleles, including pNTFlagUBR1, pNTFlagUBR1MB6, pNTFlagUBR1MR1 or pNTFlagUBR1MR2 were cultured in raffinose medium to an OD<sub>600</sub> of 0.8–1.0, then switched to galactose medium for induction of Ubc2p expression for 6 h. Cells were harvested and resuspended in DB buffer (50 mM NaCl, 1 mM EDTA, 50 mM Na-HEPES pH 7.5) containing 1× protease inhibitor mix (Boehringer Mannheim), and thereafter lysed by vortexing with glass beads (Bartel *et al.*, 1990). For the co-immunoprecipitation of Ubc2p with Ubr1p (Figure 2B, top panel), supernatants were incubated with anti-HA antiserum (Babco, Berkeley, CA), followed by SDS-8% PAGE and immunoblotting with a monoclonal anti-FLAG antibody (Sigma, St Louis, MO). The expression levels of Ubr1p and Ubc2p in these transformants were monitored by immunoblotting with anti-FLAG antibody (Figure 2B, middle panel) and anti-HA antibody (Figure 2B, bottom panel), respectively. SDS-8% PAGE was used.

#### Metabolic labeling and immunoprecipitation

*Saccharomyces cerevisiae* cells from 10 ml cultures (OD<sub>600</sub> 0.8–1.2) in SD media containing 0.1 mM CuSO<sub>4</sub> were labeled for 30 min at 30°C with 0.15 mCi of [<sup>35</sup>S]methionine/cysteine (EXPRESS, New England Nuclear, Boston, MA). The same conditions were used to label FLAG-Ubr1p and its derivatives for the GST pull-down assays, except that no CuSO<sub>4</sub> was added. To assay the *in vivo* ubiquitylation of N-end rule substrates, cells were labeled for 6 min, followed by the addition of 2 M NEM in dimethylformamide to a final concentration of 50 mM (Johnsson and Varshavsky, 1994). Two methods were used to prepare cell extracts. For the GST pull-down assays, cell extracts were prepared as described above for co-immunoprecipitation immunoblots. For the experiments shown in Figures 3C, 4B, 5A and 6, labeled cells were pelleted, and resuspended in 0.8 ml of buffer A (1% Triton X-100, 0.15 M NaCl, 1 mM EDTA, 50 mM Na-HEPES pH 7.5) containing protease inhibitors, and were lysed as above. The extracts were centrifuged at 12 000 g for 10 min, and supernatants containing equal amounts of CCl<sub>3</sub>COOH-insoluble <sup>35</sup>S were used for the GST assays described above or for immunoprecipitations. To detect Ubr1p (Figures 2A and 5A, lower panels), immunoprecipitations were carried out with a monoclonal anti-FLAG antibody conjugated to agarose beads (Kodak, Rochester, NY), followed by SDS-8% PAGE. For the co-immunoprecipitation of Leu- $\beta$ gal with Ubr1p (Figure 5A, top panel), supernatants were incubated with a monoclonal anti- $\beta$ gal antibody (Promega, Madison, WI), followed by SDS-8% PAGE. For measurements of X- $\beta$ gal stability using the

UPR technique (Figures 3C, 4B and 6), a mixture of an anti-HA antiserum (Babco, Berkeley, CA) and the monoclonal anti- $\beta$ gal antibody was incubated with  $^{35}$ S-labeled cell extracts. Immunoprecipitated proteins were separated by SDS-12% PAGE. For detection of the *in vivo* ubiquitylation, cells labeled for 6 min were lysed using the SDS-boiling method (Hochstrasser and Varshavsky, 1990). Briefly, a pellet of NEM-treated cells (see above) was mixed with an equal volume of the 2× SDS buffer at 100°C (2% SDS, 30 mM dithiothreitol, 90 mM Na-HEPES pH 7.5) plus 50 mM NEM, and incubated at 100°C for 5 min. The extracts were diluted 10-fold with buffer A plus protease inhibitors and 50 mM NEM. CCl<sub>3</sub>COOH-insoluble  $^{35}$ S was determined, and samples containing equal amounts of  $^{35}$ S were immunoprecipitated with the monoclonal anti- $\beta$ gal antibody, followed by SDS-6% PAGE.

## Acknowledgements

We are grateful to F.Du and A.Webster for the GST-Ubc2p fusion protein and the pNTFlagUBR1 and pUPR plasmids. We thank E.Craig and P.James for their two-hybrid constructs, and members of the Varshavsky laboratory, especially F.Du, H.Rao, H.-R.Wang, J.Sheng and G.Turner, for helpful discussions and comments on the manuscript. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530). Y.X. was supported in part by a postdoctoral fellowship from the NIH.

## References

- Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.
- Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.
- Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase: A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.
- Balzi,E., Choder,M., Chen,W., Varshavsky,A. and Goffeau,A. (1990) Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**, 7464–7471.
- Bartel,B. (1990) Molecular genetics of the ubiquitin system: the ubiquitin fusion proteins and proteolytic targeting mechanisms. PhD thesis. Department of Biology, MIT, Cambridge, MA.
- Bartel,B., Wünnig,I. and Varshavsky,A. (1990) The recognition component of the N-end rule pathway. *EMBO J.*, **9**, 3179–3189.
- Baumeister,W., Walz,J., Zühl,F. and Seemüller,E. (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **92**, 367–380.
- Bordallo,J., Plemper,R.K., Finger,A. and Wolf,D.H. (1998) Der3p/Hrd1p is required for endoplasmic reticulum-associated degradation of misfolded luminal and integral membrane proteins. *Mol. Biol. Cell*, **9**, 209–222.
- Borden,K.L. and Freemont,P.S. (1996) The RING finger domain: a recent example of a sequence-specific family. *Curr. Opin. Struct. Biol.*, **6**, 395–401.
- Byrd,C., Turner,G.C. and Varshavsky,A. (1998) The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.*, **17**, 269–277.
- Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.
- Chen,D.-C., Yang,B.-C. and Kuo,T.-T. (1992) One-step transformation of yeast in stationary phase. *Curr. Genet.*, **21**, 83–84.
- Dohmen,R.J., Madura,K., Bartel,B. and Varshavsky,A. (1991) The N-end rule is mediated by the UBC2 (RAD6) ubiquitin-conjugating enzyme. *Proc. Natl Acad. Sci. USA*, **88**, 7351–7355.
- Durfee,T., Becherer,K., Chen,P.L., Yeh,S.H., Yang,Y., Kilburn,A.E., Lee,W.H. and Elledge,S.J. (1993) The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev.*, **7**, 555–569.
- Elenbaas,B., Dobbelstein,M., Roth,J., Shenk,T. and Levine,A.J. (1996) The MDM2 oncoprotein binds specifically to RNA through its RING finger domain. *Mol. Med.*, **2**, 439–451.
- Fields,S. and Song,O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245–246.
- Ghislain,M., Dohmen,R.J., Levy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *EMBO J.*, **15**, 4884–4899.
- Gonda,D.K., Bachmair,A., Wünnig,I., Tobias,J.W. and Varshavsky,A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.*, **264**, 16700–16712.
- Grigoryev,S., Stewart,A.E., Kwon,Y.T., Arfin,S.M., Bradshaw,R.A., Jenkins,N.A., Copeland,N.G. and Varshavsky,A. (1996) A mouse amidase specific for N-terminal asparagine. The gene, the enzyme and their function in the N-end rule pathway. *J. Biol. Chem.*, **271**, 28521–28532.
- Hershko,A. and Ciechanover,A. (1998) The ubiquitin system. *Annu. Rev. Biochem.*, **76**, 425–479.
- Hicke,L. (1997) Ubiquitin-dependent internalization and down-regulation of plasma membrane proteins. *FASEB J.*, **11**, 1215–1226.
- Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.
- Hochstrasser,M. and Varshavsky,A. (1990) *In vivo* degradation of a transcriptional regulator: the yeast  $\alpha$ 2 repressor. *Cell*, **61**, 697–708.
- Honda,R., Tanaka,H. and Yasuda,H. (1997) Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.*, **420**, 25–27.
- Hu,G. and Fearon,E.R. (1999) Siah-1 N-terminal RING domain is required for proteolysis function and C-terminal sequences regulate oligomerization and binding to target proteins. *Mol. Cell. Biol.*, **19**, 724–732.
- Hubregtsse,J.M., Maki,C.G. and Howley,P.M. (1998) Ubiquitination of the p53 tumor suppressor. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 323–343.
- Inouye,C., Dhillon,N. and Thorner,J. (1997) Ste5 RING-H2 domain: role in Ste4 promoted oligomerization for yeast pheromone signaling. *Science*, **278**, 103–106.
- James,P., Halladay,J. and Craig,E. (1996) Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics*, **144**, 1425–1436.
- Jentsch,S., McGrath,J.P. and Varshavsky,A. (1987) The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature*, **329**, 131–134.
- Johnson,E.S., Ma,P.C., Ota,I.M. and Varshavsky,A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.*, **270**, 17442–17456.
- Johnsson,N. and Varshavsky,A. (1994) Ubiquitin-assisted dissection of protein transport across cell membranes. *EMBO J.*, **13**, 2686–2698.
- Kamura,K. et al. (1999) Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science*, **284**, 657–661.
- Koegl,M., Hoppe,T., Schlenker,S., Ulrich,H.D., Mayer,T.U. and Jentsch,S. (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell*, **96**, 635–644.
- Kwon,Y.T. et al. (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **95**, 7898–7903.
- Kwon,Y.T., Kashina,A.S. and Varshavsky,A. (1999a) Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. *Mol. Cell. Biol.*, **19**, 182–193.
- Kwon,Y.T., Lévy,F. and Varshavsky,A. (1999b) Bivalent inhibitor of the N-end rule pathway. *J. Biol. Chem.*, **274**, 18135–18139.
- Lévy,F., Johnsson,N., Rumenapf,T. and Varshavsky,A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl Acad. Sci. USA*, **93**, 4907–4912.
- Lévy,F., Johnston,J.A. and Varshavsky,A. (1999) Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur. J. Biochem.*, **259**, 244–252.
- Madura,K. and Varshavsky,A. (1994) Degradation of Ga by the N-end rule pathway. *Science*, **265**, 1454–1458.
- Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.
- Mumberg,D., Muller,R. and Funk,M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*—comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.*, **22**, 5767–5768.
- Nuber,U. and Scheffner,M. (1999) Identification of determinants in E2 ubiquitin conjugating enzymes required for hec3E3 ubiquitin-protein ligase interactions. *J. Biol. Chem.*, **274**, 7576–7582.
- Ohta,T., Michel,J.J., Schottelius,A.J. and Xiong,Y. (1999) Roc1, a homolog of Apc11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol. Cell*, **3**, 535–541.

- Peters,J.-M., Harris,J.R. and Finley,D. (eds) (1998) *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY.
- Pickart,C.M. (1997) Targeting of substrates to 26S proteasome. *FASEB J.*, **11**, 1055–1066.
- Reiss,Y., Kaim,D. and Hershko,A. (1988) Specificity of binding of N-terminal residues of proteins to ubiquitin–protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.*, **263**, 2693–2698.
- Saurin,A.J., Borden,K.L.B., Boddy,M.N. and Freemont,P.S. (1996) Does this have a familiar RING? *Trends Biochem. Sci.*, **21**, 208–214.
- Scheffner,M., Huibregtse,J.M., Vierstra,R.D. and Howley,P.M. (1993) The HPV-16 E6 and E6-AP complex functions as a ubiquitin–protein ligase in the ubiquitination of p53. *Cell*, **75**, 495–505.
- Scheffner,M., Huber,U. and Huibregtse,J.M. (1995) Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature*, **373**, 81–83.
- Scheffner,M., Smith,S. and Jentsch,S. (1998) The ubiquitin conjugation system. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 65–98.
- Seol,J.H. et al. (1999) Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev.*, **13**, 1614–1626.
- Sherman,F. (1991) Getting started with yeast. *Methods Enzymol.*, **194**, 3–21.
- Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. *Genetics*, **122**, 19–27.
- Skowrya,D., Koeppp,D.M., Kamura,T., Conrad,M.N., Conaway,R.C., Conaway,J.W., Elledge,S.J. and Harper,J.W. (1999) Reconstitution of G<sub>1</sub> cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science*, **284**, 662–665.
- Stewart,A.E., Arfin,S.M. and Bradshaw,R.A. (1995) The sequence of porcine protein NH<sub>2</sub>-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 25–28.
- Tan,P., Fuchs,S.Y., Chen,A., Wu,K., Gomez,C., Ronai,Z. and Pan,Z.-Q. (1999) Recruitment of Roc1–Cul1 ubiquitin ligase by Skp1 and HOS to catalyze the ubiquitination of IκBα. *Mol. Cell*, **3**, 527–533.
- Tyers,M. and Willemse,A.R. (1999) One ring to rule a superfamily of E3 ubiquitin ligases. *Science*, **284**, 602–604.
- Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.
- Varshavsky,A. (1997) The ubiquitin system. *Trends Biochem. Sci.*, **22**, 383–387.
- Wang,G., Yang,J. and Huibregtse,J.M. (1999) Functional domains of the Rsp5 ubiquitin–protein ligase. *Mol. Cell. Biol.*, **19**, 342–352.
- Watkins,J.F., Sung,P., Prakash,S. and Prakash,L. (1993) The extremely conserved amino-terminus of RAD6 ubiquitin-conjugating enzyme is essential for N-end rule-dependent protein degradation. *Genes Dev.*, **7**, 250–261.
- Wilkinson,K. and Hochstrasser,M. (1998) The deubiquitinating enzymes. In Peters,J.M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 99–125.
- Worthy lake,D.K., Prakash,S., Prakash,L. and Hill,C.P. (1998) Crystal structure of the *Saccharomyces cerevisiae* ubiquitin-conjugating enzyme Rad6 at 2.6 Å resolution. *J. Biol. Chem.*, **273**, 6271–6276.
- Yu,H., Peters,J.M., King,R.W., Page,A.M., Hieter,P. and Kirschner,M.W. (1998) Identification of a cullin homology region in a subunit of the anaphase-promoting complex. *Science*, **279**, 1219–1222.
- Zachariae,W., Shevchenko,A., Andrews,P.D., Ciosk,R., Galova,M., Stark,J.R., Mann,M. and Nasmyth,K. (1998) Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins. *Science*, **279**, 1216–1219.

Received July 30, 1999; revised and accepted October 8, 1999

## Bivalent Inhibitor of the N-end Rule Pathway\*

(Received for publication, February 9, 1999)

Yong Tae Kwon‡, Frédéric Lévy§, and Alexander Varshavsky¶||

From the ‡Division of Biology, California Institute of Technology, Pasadena, California 91125  
and the §Ludwig Institute for Cancer Research, 155, ch. Des Boveresses, CH-1066 Epalinges, Switzerland

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Ubr1p, the recognition (E3) component of the *Saccharomyces cerevisiae* N-end rule pathway, contains at least two substrate-binding sites. The type 1 site is specific for N-terminal basic residues Arg, Lys, and His. The type 2 site is specific for N-terminal bulky hydrophobic residues Phe, Leu, Trp, Tyr, and Ile. Previous work has shown that dipeptides bearing either type 1 or type 2 N-terminal residues act as weak but specific inhibitors of the N-end rule pathway. We took advantage of the two-site architecture of Ubr1p to explore the feasibility of bivalent N-end rule inhibitors, whose expected higher efficacy would result from higher affinity of the cooperative (bivalent) binding to Ubr1p. The inhibitor comprised mixed tetramers of  $\beta$ -galactosidase that bore both N-terminal Arg (type 1 residue) and N-terminal Leu (type 2 residue) but that were resistant to proteolysis *in vivo*. Expression of these constructs in *S. cerevisiae* inhibited the N-end rule pathway much more strongly than the expression of otherwise identical  $\beta$ -galactosidase tetramers whose N-terminal residues were exclusively Arg or exclusively Leu. In addition to demonstrating spatial proximity between the type 1 and type 2 substrate-binding sites of Ubr1p, these results provide a route to high affinity inhibitors of the N-end rule pathway.

Among the targets of the N-end rule pathway are intracellular proteins bearing destabilizing N-terminal residues (1, 2). This proteolytic pathway is one of several pathways of the ubiquitin (Ub)<sup>1</sup> system, whose diverse functions include the regulation of cell growth, division, differentiation, and responses to stress (3–6). Ub is a 76-residue eukaryotic protein that exists in cells either free or conjugated to other proteins. Many of the Ub-dependent regulatory circuits involve processive degradation of ubiquitylated proteins by the 26 S proteasome, an ATP-dependent multisubunit protease (7, 8).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase) to Arg, one of the primary destabilizing

residues (reviewed in Refs. 1 and 9). In mammals, two distinct N-terminal amidases specific, respectively, for N-terminal Asn or Gln mediate the conversion of these tertiary destabilizing residues into the secondary destabilizing residues Asp or Glu (10, 11). The set of secondary destabilizing residues in vertebrates contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (9, 12, 13).

The primary destabilizing N-terminal residues are bound directly by N-recognin, the E3 (recognition) component of the N-end rule pathway. In *S. cerevisiae*, N-recognin is the *UBR1*-encoded 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues: Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (1, 14). The *Ubr1* genes encoding mouse and human N-recognin (also called E3 $\alpha$ ) have been cloned as well (15). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (1, 12, 16, 17). N-recognin can also target short-lived proteins such as Cup9p (18) and Gpa1p (19, 20), which lack destabilizing N-terminal residues. The Ubr1p-recognized degradation signals of these proteins remain to be characterized in detail.

The known functions of the N-end rule pathway include the control of di- and tripeptide import in *S. cerevisiae* through the degradation of Cup9p, a transcriptional repressor of the peptide transporter gene *PTR2* (18, 21); a mechanistically undefined role in the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (22); the degradation of Gpa1p, a G $\alpha$  protein of *S. cerevisiae* (19, 20); and the conditional degradation of alphavirus RNA polymerase in virus-infected metazoan cells (23). Physiological N-end rule substrates were also identified among the proteins secreted into the cytosol of the host cell by intracellular parasites such as the bacterium *Listeria monocytogenes* (24). Short half-lives of these proteins are required for the efficient presentation of their peptides to the immune system (24). A partial inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (25) and to delay limb regeneration in amphibians (26). Recent evidence suggests that the N-end rule pathway mediates a large fraction of the muscle protein turnover (27) and plays a role in catabolic states that result in muscle atrophy (28).

Targeted mutagenesis has been used to inactivate the N-end rule pathway in *Escherichia coli* and *S. cerevisiae* (14, 29). Analogous mutants have recently been constructed in the mouse as well.<sup>2</sup> These approaches notwithstanding, an efficacious inhibitor of the N-end rule pathway would be useful as well, especially with organisms less tractable genetically. The emerging understanding of the N-end rule pathway in mammals suggests that selective inhibition or activation of this proteolytic system may also have medical applications. Previ-

\* This work was supported by National Institutes of Health Grant DK39520 (to A. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Div. of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125. Tel.: 626-395-3785; Fax: 626-440-9821; E-mail: varsh@its.caltech.edu.

‡ The abbreviations used are: Ub, ubiquitin;  $\beta$ gal, *E. coli*  $\beta$ -galactosidase; E3, ubiquitin-protein ligase; ha, hemagglutinin.

<sup>2</sup> Y. T. Kwon and A. Varshavsky, unpublished data.

ous work has shown that millimolar concentrations of amino acid derivatives such as dipeptides bearing destabilizing N-terminal residues can selectively inhibit the N-end rule pathway in extracts from rabbit reticulocytes (12, 17) and *Xenopus* eggs (13), and in intact *S. cerevisiae* cells as well (16). However, the same dipeptides were observed to have at most marginal effects on the N-end rule pathway in intact mammalian cells.<sup>3</sup> One limitation of dipeptide inhibitors is their apparently low affinity for the type 1 and the type 2 site of N-recognin (30).

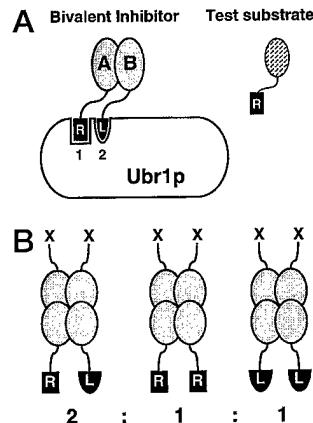
In the present work, we explored the possibility that a bivalent ligand can bind simultaneously to the type 1 and type 2 sites of N-recognin (see Fig. 1A). Similarly to the previously characterized bivalent interactions that involve either macromolecules or small molecules (31, 32), the cooperativity of binding at two independent, mutually nonexclusive sites would be expected to increase the affinity between N-recognin and a bivalent inhibitor by orders of magnitude, in comparison with the affinity of a monovalent binding by the same compound. We show that a bivalent inhibitor of the N-end rule pathway is feasible and consider the implications of this advance.

#### EXPERIMENTAL PROCEDURES

**Strains and General Techniques**—The *S. cerevisiae* strains used were JD52 (*MATa* *ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801*) and JD55 (*MATa* *ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801 ubr1Δ::HIS3*) (19, 33). Cells were grown on rich (YPD) or synthetic medium containing either 2% dextrose (SD medium), 2% galactose (SG medium), or 2% raffinose (SR medium) (34). To induce the *P<sub>CUP1</sub>* promoter, CuSO<sub>4</sub> was added to a final concentration of 0.1 mM. Transformation of *S. cerevisiae* was carried out using the lithium acetate method (35).

**Plasmids**—The high copy (2 $\mu$ -based) plasmids pRΔ-βgal-TRP1 and pRΔ-βgal-HIS3, which expressed Arg-e<sup>AK</sup>-βgal (Ub-Arg-e<sup>AK</sup>-βgal) (see Fig. 2A) from the galactose-inducible *P<sub>CYC1/GAL4</sub>* hybrid promoter (2), were produced by replacing the *URA3* marker gene of pFL7 with either *TRP1* or *HIS3*. pLΔ-βgal-TRP1 and pLΔ-βgal-HIS3, both of which expressed Leu-e<sup>AK</sup>-βgal (Ub-Leu-e<sup>AK</sup>-βgal), were produced by replacing the Ub-Arg domain of pRΔ-βgal-TRP1 and pRΔ-βgal-HIS3 with Ub-Leu domain of the pLL2 plasmid.<sup>4</sup> The plasmid pFL7 was produced from pUB23-R (2) by converting the lysine codons 15 and 17 of the extension e<sup>K</sup> into arginine codons (36, 37), yielding a construct encoding the extension e<sup>AK</sup> in front of a βgal moiety lacking the first 23 residues of wild type βgal (see Fig. 2A). The low copy, pRS315 vector-derived (38) plasmid pR-eΔKhaUra3-R3R7 expressed Arg-e<sup>AK</sup>-ha-Ura3p<sup>K3R,K7R</sup> (Ub-Arg-e<sup>AK</sup>-ha-Ura3p<sup>K3R,K7R</sup>) from the *P<sub>CUP1</sub>* promoter. Arg-e<sup>AK</sup>-ha-Ura3p<sup>K3R,K7R</sup> (see Fig. 2B) is called Arg-Ura3p in the main text. In this N-end rule substrate, the residues Lys-3 and Lys-7 of the *S. cerevisiae* Ura3p were converted to arginines (see "Results and Discussion"). In addition, the ha epitope tag (39) was placed between e<sup>AK</sup> and Ura3p<sup>K3R,K7R</sup> (see Fig. 2B). The plasmid pR-eΔKhaUra3-R3R7 was produced from pFL1 (encoding Ub-Arg-e<sup>AK</sup>-ha-Ura3p) through site-directed mutagenesis of the *URA3* codons for Lys-3 and Lys-7. pFL1 was produced from pKM1235 (which encoded Ub-Arg-e<sup>K</sup>-ha-Ura3p)<sup>5</sup> by converting the e<sup>K</sup>-coding sequence into the one encoding e<sup>AK</sup>.

**Pulse-Chase and Plating Efficiency Assays**—Pulse-chase assays with *S. cerevisiae* in mid-exponential growth ( $A_{600}$  of ~1) utilized <sup>35</sup>S-EXPRESS (NEN Life Science Products) and were carried out as described previously (10, 19), including the immunoprecipitation with anti-βgal and anti-ha antibodies and quantitation with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). To determine plating efficiency, *S. cerevisiae* strains JD52 (*UBR1*) and JD55 (*ubr1Δ*) expressing Arg-Ura3p (Ub-Arg-e<sup>AK</sup>-ha-Ura3p<sup>K3R,K7R</sup>; see Fig. 2B) were co-transformed with plasmids indicated in the legend to Fig. 3. The transformants were cultured in the raffinose-based medium (SR) lacking Leu, His, and Trp for 20 h. The cultures were then diluted into the otherwise identical galactose-containing (SG) medium to a final  $A_{600}$  of 0.1. At an  $A_{600}$  of 0.4, cultures were either supplemented with 0.1 mM CuSO<sub>4</sub> or left unsupplemented. At the  $A_{600}$  of 1.0, the cultures were diluted with SG (which lacks Leu, His, and Trp) either containing or lacking 0.1 mM CuSO<sub>4</sub> and were plated on the plates of the same medium composition that also either contained or lacked uracil. The plating efficiency (%)



**FIG. 1. The concept of a bivalent inhibitor of the N-end rule pathway.** **A**, the type 1 and type 2 sites of *S. cerevisiae* Ubr1p (N-recognin), which are specific, respectively, for the basic (Arg, Lys, and His) and bulky hydrophobic (Phe, Leu, Trp, Tyr, and Ile) N-terminal residues. In the diagram, the type 1 and type 2 sites are occupied by their ligands, the N-terminal Arg and Leu, borne by a heterodimeric bivalent inhibitor (actually, a tetrameric βgal-based protein in the present work). A test substrate bearing Arg, a type 1 destabilizing N-terminal residue is shown as well. The test substrate, in contrast to the protein-based inhibitor, bears at least one internal Lys residue (not indicated in the diagram) that can function as a component of the N-degron. The type 1 and type 2 sites of N-recognin are shown located close together in the N-terminal region of the 225-kDa Ubr1p. The recent genetic dissection of the Ubr1p substrate-binding sites<sup>6</sup> placed the type 1 and type 2 sites close together in the ~60-kDa N-terminal region of the 225-kDa Ubr1p. **B**, a diagram illustrating the expected frequencies of heterodimeric (Arg- and Leu-bearing) dimers within a βgal-based bivalent inhibitor. Specifically, at equal levels of expression of the two βgal-based polypeptide chains, 50% of βgal tetramers would be expected to be heterotetramers in which at least one of the two dimers bears different (Arg and Leu) N-terminal residues. In the βgal tetramer, the two N termini of each dimer are spatially close, exposed, and oriented in the same direction (40). See also "Results and Discussion."

was defined as the ratio of the number of colonies on SG (-Leu, -His, -Trp, -Ura) plates to the number of colonies on SG (-Leu, -His, -Trp) plates, at the same concentration of CuSO<sub>4</sub>. For each measurement, colonies on 15 plates were counted to yield the average number of colonies per plate.

#### RESULTS AND DISCUSSION

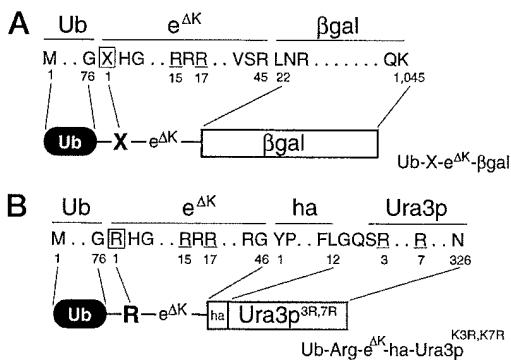
We constructed a bivalent N-end rule inhibitor (Fig. 1A) from the previously studied N-end rule substrates derived from *E. coli* βgal (2). In eukaryotes, linear Ub-protein fusions are rapidly cleaved by deubiquitylating enzymes at the Ub-protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues, a technical advance that led to the finding of the N-end rule (2). A βgal-based N-end rule substrate contains a destabilizing N-terminal residue (produced *in vivo* using the Ub fusion technique (1)); a ~45-residue, *E. coli* Lac repressor-derived N-terminal extension called e<sup>K</sup> (extension e bearing lysines K); and the βgal moiety lacking its first 21 residues. The resulting X-e<sup>AK</sup>-βgal is a short-lived protein in both yeast and mammalian cells, whereas an otherwise identical protein bearing a stabilizing N-terminal residue such as Met or Val is metabolically stable (1, 2). An N-degron comprises a destabilizing N-terminal residue and a Lys residue (or residues), the latter being the site of formation of a multi-Ub chain (1, 36). (Ubr1p can also recognize a set of other, internal degrons, which remain to be characterized (18).) If Lys-15 and Lys-17 of the e<sup>K</sup> extension are replaced by the Arg residues (which cannot be ubiquitylated), the resulting X-e<sup>AK</sup>-βgal (Fig. 2A) is long-lived *in vivo* even if its N-terminal residue is destabilizing in the N-end rule (1, 37).

In the present work, we used the metabolically stable Arg-

<sup>3</sup> F. Lévy and A. Varshavsky, unpublished data.

<sup>4</sup> M. Ghislain and A. Varshavsky, unpublished data.

<sup>5</sup> K. Madura and A. Varshavsky, unpublished data.

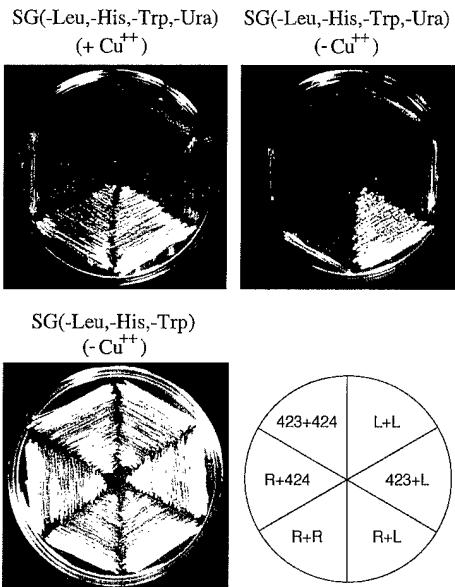


**FIG. 2. Designs of bivalent inhibitor and test substrate.** *A*, the  $\beta\text{gal}$ -based fusions (the residue  $X$  was either Arg or Leu) used to construct the Arg/Leu-bearing bivalent inhibitor. The Ub moiety of the fusions was cotranslationally removed *in vivo* by deubiquitinating enzymes (1). The ~45-residue, *E. coli* Lac repressor-derived sequence termed  $e^{\Delta K}$  (extension ( $e$ ) lacking lysines ( $\Delta K$ )), is described in the main text. The  $\beta\text{gal}$  part of the fusion lacked the first 21 residues of wild type  $\beta\text{gal}$  (2). *B*, the Ura3p-based N-end rule substrate, Arg- $e^{\Delta K}$ -ha-Ura3p<sup>K3R,K7R</sup>, derived from Ub-Arg- $e^{\Delta K}$ -ha-Ura3p<sup>K3R,K7R</sup> and denoted Arg-Ura3p, is described in the main text.

$e^{\Delta K}$ - $\beta\text{gal}$  (produced from Ub-Arg- $e^{\Delta K}$ - $\beta\text{gal}$ ) and Leu- $e^{\Delta K}$ - $\beta\text{gal}$  (produced from Ub-Leu- $e^{\Delta K}$ - $\beta\text{gal}$ ). These proteins retain the ability to bind, respectively, to the type 1 and type 2 sites of N-recognin but cannot be ubiquitylated (37), apparently because the most N-terminal Lys residue in X- $e^{\Delta K}$ - $\beta\text{gal}$ , at position 239, is too far from the N terminus of the protein. In the  $\beta\text{gal}$  tetramer, the two N termini of each dimer are spatially close, exposed, and oriented in the same direction (40). At equal levels of expression of the two  $\beta\text{gal}$ -based polypeptide chains such as Arg- $e^{\Delta K}$ - $\beta\text{gal}$  and Leu- $e^{\Delta K}$ - $\beta\text{gal}$ , 50% of  $\beta\text{gal}$  trimers would be expected to be heterotetramers in which at least one of the two dimers bears different (Arg and Leu) N-terminal residues (Fig. 1*B*). If the type 1 and type 2 substrate-binding sites of the 225-kDa Ubr1p are appropriately located and oriented, they might be able to bind the Arg- and Leu-bearing subunits of the mixed  $\beta\text{gal}$  tetramer, especially in view of the presumed flexibility of the  $e^{\Delta K}$  extension (1) (Fig. 1*A*).

The reporter N-end rule substrate in this study was Arg- $e^{\Delta K}$ -ha-Ura3p<sup>K3R,K7R</sup>, denoted below as Arg-Ura3p (Fig. 2*B*). This ha-tagged, type 1 N-end rule substrate was produced from Ub-Arg- $e^{\Delta K}$ -ha-Ura3p<sup>K3R,K7R</sup> through the cotranslational *in vivo* cleavage by deubiquitinating enzymes (1, 6, 41). The lysine-lacking  $e^{\Delta K}$  extension of Arg- $e^{\Delta K}$ -ha-Ura3p<sup>K3R,K7R</sup>, and the replacement of the first two lysines of the Ura3p moiety with arginines were used to decrease the rate of degradation of Arg-Ura3p by the N-end rule pathway and also to reduce the slow but detectable degradation of Arg-Ura3p by yet another pathway, through a degron distinct from the N-degron.<sup>3</sup> Several Lys residues of Ura3p other than Lys-3 and Lys-7 are also close to its N terminus, thus accounting for the absence, in this case, of the all-or-none effect on the reporter degradation that is observed when  $e^K$  is replaced with  $e^{\Delta K}$  in an X- $e^K$ - $\beta\text{gal}$  substrate (37). The Lys-3 → Arg and Lys-7 → Arg modifications decreased the enzymatic activity of the Ura3p moiety.<sup>2</sup> The reduced enzymatic activity of Ura3p<sup>K3R,K7R</sup> facilitated selection assays (Figs. 3 and 4).

The first bivalent inhibitor assay employed *ura3* *S. cerevisiae* expressing Arg-Ura3p (Fig. 2*B*) from the uninduced  $P_{CUP1}$  promoter. The Ubr1p-mediated degradation of Arg-Ura3p ( $t_{1/2}$  of ~8 min) and its correspondingly low steady-state concentration rendered wild type (*UBR1*) cells phenotypically  $Ura^-$ , whereas *ubr1Δ* strains expressing Arg-Ura3p were phenotypically  $Ura^+$  (Figs. 3 and 4 and data not shown). Cells expressing Arg-Ura3p were cotransformed with two control plasmids (vectors;

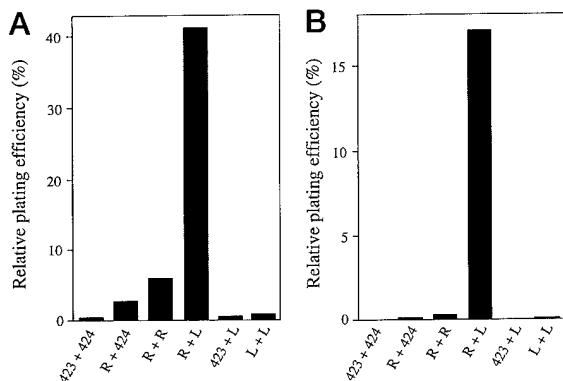


**FIG. 3. A bivalent, but not monovalent, inhibitor of Ubr1p confers  $Ura^+$  phenotype on cells expressing the short-lived N-end rule substrate Arg-Ura3p.** *S. cerevisiae* JD52 (*UBR1*) expressing Arg-Ura3p (Ub-Arg-Ura3p), a short-lived reporter (Fig. 2*B*), were cotransformed, alternatively, with pRS423 (*HIS3*-based control vector) and pRS424 ( $P_{GAL1}$ , *TRP1*-based control vector) (denoted as 423+424); with pRS424 and pRA- $\beta\text{gal}$ -HIS3, expressing Arg- $e^{\Delta K}$ - $\beta\text{gal}$  (Ub-Arg- $e^{\Delta K}$ - $\beta\text{gal}$ ) (denoted as R+424; monovalent inhibitor); with pRS423 and pLA- $\beta\text{gal}$ -TRP1, expressing Leu- $e^{\Delta K}$ - $\beta\text{gal}$  (Ub-Leu- $e^{\Delta K}$ - $\beta\text{gal}$ ) (denoted as 423+L); with pRA- $\beta\text{gal}$ -HIS3 and pRA- $\beta\text{gal}$ -TRP1, both expressing Arg- $e^{\Delta K}$ - $\beta\text{gal}$  (denoted as R+R; monovalent inhibitor); with pLA- $\beta\text{gal}$ -HIS3 and pLA- $\beta\text{gal}$ -TRP1, both expressing Leu- $e^{\Delta K}$ - $\beta\text{gal}$  (denoted as L+L); or with pRA- $\beta\text{gal}$ -HIS3 and pLA- $\beta\text{gal}$ -TRP1, expressing Arg- $e^{\Delta K}$ - $\beta\text{gal}$  and Leu- $e^{\Delta K}$ - $\beta\text{gal}$  (denoted as R+L; the bivalent inhibitor). Cells were streaked on SG medium containing 0.1 mM CuSO<sub>4</sub> and lacking Leu, His, Trp, and Ura (upper left panel), on the otherwise identical medium lacking the added CuSO<sub>4</sub> (upper right panel), or on the Ura-containing SG medium lacking Leu, His, and Trp (controls; lower left panel). Plates were incubated at 30 °C for 3 days.

423+424 in Fig. 3). Alternatively, these cells were cotransformed with two plasmids (bearing different selectable markers) that expressed either Arg- $e^{\Delta K}$ - $\beta\text{gal}$  alone (R+R in Fig. 3), Leu- $e^{\Delta K}$ - $\beta\text{gal}$  alone (L+L in Fig. 3), or both of them together (R+L in Fig. 3; the bivalent inhibitor mode) from a galactose-inducible promoter. Pairs of alternatively marked plasmids were used to make certain that the conditions of expression and the total amounts of  $\beta\text{gal}$ -based proteins produced remained the same in all of these settings. The transformants were streaked on SG medium lacking uracil.

Remarkably, only those Arg-Ura3p-expressing cells that expressed both Arg- $e^{\Delta K}$ - $\beta\text{gal}$  and Leu- $e^{\Delta K}$ - $\beta\text{gal}$  became  $Ura^+$  under these conditions (Fig. 3). The cells that expressed either Arg- $e^{\Delta K}$ - $\beta\text{gal}$  alone or Leu- $e^{\Delta K}$ - $\beta\text{gal}$  alone remained  $Ura^-$ , as did the cells that received control plasmids (Fig. 3). (The same cells grew equally well in the control SG medium containing uracil (Fig. 3, bottom left panel).) Note that the monovalent inhibitors were ineffective despite the fact that the concentration of either the Arg-based N terminus alone or the Leu-based N terminus alone was twice the concentration of the same N termini in the case of the bivalent inhibitor.

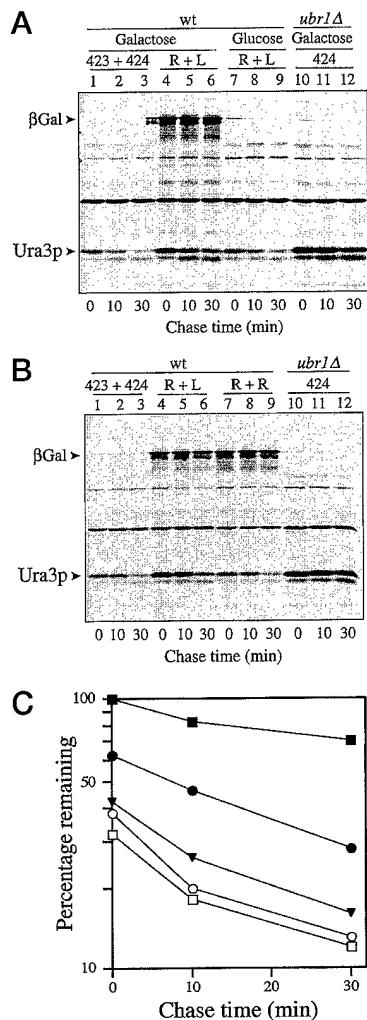
To quantify the effect of coexpressing Arg- $e^{\Delta K}$ - $\beta\text{gal}$  and Leu- $e^{\Delta K}$ - $\beta\text{gal}$  on the rescue of the  $Ura^+$  phenotype, a plating efficiency assay was carried out with the same transformants. Equal amounts of cells were plated on SG(+Ura) and SG(-Ura) plates, and the numbers of colonies were determined. When the Arg-Ura3p reporter was expressed at a sufficiently low rate (uninduced  $P_{CUP1}$  promoter), cells became



**FIG. 4. Plating efficiencies of *S. cerevisiae* expressing Arg-Ura3p in the presence of bivalent and monovalent inhibitors of the N-end rule pathway.** *A*, *S. cerevisiae* JD52 (*UBR1*) and JD55 (*ubr1Δ*) expressing Arg-Ura3p were cotransformed with the sets of plasmids described and denoted in the legend to Fig. 3. The transformants were cultured as described under “Experimental Procedures” and plated on either SG(-Leu, -His, -Trp, -Ura) plates or control plates SG(-Leu, -His, -Trp) containing 0.1 mM CuSO<sub>4</sub>. The plating efficiencies shown are the values produced by normalization against the absolute plating efficiency (92%) of the positive control: the *ubr1Δ* strain JD55 expressing Arg-Ura3p and bearing the vector pRS424. *B*, the same experiment was done using plates lacking the added CuSO<sub>4</sub>. The plating efficiencies shown are the values produced by normalization against the positive control used in *A*. Under these growth conditions (no added CuSO<sub>4</sub>), the absolute plating efficiency of the positive control was 26%.

Ura<sup>+</sup> (through metabolic stabilization of Arg-Ura3p) only in the presence of both Arg-e<sup>ΔK</sup>-βgal and Leu-e<sup>ΔK</sup>-βgal (Fig. 4B). A weak stabilizing effect of Arg-e<sup>ΔK</sup>-βgal alone could be detected only at a ~20-fold higher level of Arg-Ura3p expression (induced P<sub>CUP1</sub> promoter) (Fig. 4A). No stabilization of Arg-Ura3p was observed in the presence of Leu-e<sup>ΔK</sup>-βgal under any conditions (Fig. 4), confirming the specificity of inhibition in regard to the type (basic or bulky hydrophobic) of the primary destabilizing N-terminal residue of the reporter. Higher sensitivity of this assay at the higher level of Arg-Ura3p expression results from a higher steady-state level of the short-lived Arg-Ura3p, so that even its marginal stabilization suffices to render a small fraction of cells Ura<sup>+</sup> (Fig. 4A; compare with Fig. 4B).

To analyze directly the *in vivo* degradation of Arg-Ura3p in the presence of different combinations of X-e<sup>ΔK</sup>-βgal proteins, the transformants of Figs. 3 and 4 were subjected to pulse-chase analysis, with immunoprecipitation of both Arg-Ura3p and the (long-lived) X-e<sup>ΔK</sup>-βgals (Fig. 5). Quantitation of the resulting electrophoretic patterns (Fig. 5C) confirmed and extended the conclusions reached through phenotypic analyses (Figs. 3 and 4). Specifically, the normally short-lived Arg-Ura3p (Fig. 5A, lanes 1–3) was strongly (but still incompletely) stabilized in the presence of both Arg-e<sup>ΔK</sup>-βgal and Leu-e<sup>ΔK</sup>-βgal (Fig. 5A, lanes 4–6; compare with lanes 1–3 and 7–9). This stabilization was manifested especially clearly as an increase in the relative amount of Arg-Ura3p at the beginning of chase (time 0), indicating reduced degradation of Arg-Ura3p during the pulse (Fig. 5C). This latter degradation pattern, termed “zero point effect,” is caused by the previously demonstrated preferential targeting of newly formed (as distinguished from conformationally mature) protein substrates by the N-end rule pathway (16, 42). The increased steady-state level of Arg-Ura3p in the presence of both Arg-e<sup>ΔK</sup>-βgal and Leu-e<sup>ΔK</sup>-βgal accounted for the results of phenotypic analyses (Figs. 3 and 4). The much smaller but detectable stabilization of Arg-Ura3p by Arg-e<sup>ΔK</sup>-βgal alone (Fig. 5C) was consistent not only with the inability of Arg-e<sup>ΔK</sup>-βgal to confer the Ura<sup>+</sup> phenotype on cells expressing Arg-Ura3p from uninduced P<sub>CUP1</sub> promoter but also



**FIG. 5. Metabolic stabilization of Arg-Ura3p in the presence of bivalent N-end rule inhibitor.** *A*, *S. cerevisiae* JD52 (*UBR1*) expressing Arg-Ura3p, a short-lived Ura3p-based reporter (Fig. 2B), from the induced P<sub>CUP1</sub> promoter, were cotransformed, alternatively, with either pRS423 (*HIS3*-based control vector) and pRS424 (P<sub>GAL1</sub>, *TRP1*-based control vector) (denoted as 423+424), or with pRΔ-βgal-TRP1 and pLΔ-βgal-HIS3, expressing Arg-e<sup>ΔK</sup>-βgal and Leu-e<sup>ΔK</sup>-βgal (denoted as R+L; the bivalent inhibitor). Control JD55 (*ubr1Δ*) cells expressing Arg-Ura3p were transformed with pRS424. Cells grown in either dextrose-containing SD medium (no expression of βgal) or galactose-containing SG medium were labeled with [<sup>35</sup>S]methionine/cysteine for 5 min at 30 °C, followed by a chase for 0, 10, and 30 min, extraction, immunoprecipitation, and SDS-10% polyacrylamide gel electrophoresis. *B*, same as in *A*, but cells were also cotransformed with the plasmids pRΔ-βgal-TRP1 and pRΔ-βgal-HIS3, both expressing Arg-e<sup>ΔK</sup>-βgal (denoted as R+R). The assays were carried out in SG medium. *C*, *in vivo* decay curves of Arg-Ura3p (Fig. 2B) in wild type (JD52) and *ubr1Δ* (JD55) cells. The patterns in *A* and *B* were quantified as described under “Experimental Procedures.” The initial amounts of Arg-Ura3p were normalized against the amount in *ubr1Δ* cells (100%). Note that the inhibitors also altered the zero point effect (degradation of a reporter during the pulse (16, 42)). ●, degradation of Arg-Ura3p in *UBR1* (JD52) cells in the presence of both Arg-e<sup>ΔK</sup>-βgal and Leu-e<sup>ΔK</sup>-βgal (cells were grown in SG medium); ○, the same transformants were grown in SD medium where βgal fusions were not expressed; ▼, JD52 cells expressing Arg-Ura3p were transformed with the two alternatively marked plasmids expressing Arg-e<sup>ΔK</sup>-βgal and grown in SG medium; □, JD52 cells expressing Arg-Ura3p were transformed with the two alternatively marked control vectors and grown in SG medium; ■, *ubr1Δ* (JD55) cells expressing Arg-Ura3p were transformed with control vectors and grown in SG medium.

with the partial rescue of the Ura<sup>+</sup> phenotype by Arg-e<sup>ΔK</sup>-βgal in cells expressing Arg-Ura3p from the induced P<sub>CUP1</sub> (Figs. 3 and 4 and data not shown).

The Arg/Leu-e<sup>ΔK</sup>-βgal-based bivalent inhibitor of the present work, although surprisingly potent (Fig. 4B), is obviously far from optimal even for a protein-based inhibitor; because βgal is a homotetramer, only ~50% of the coexpressed Arg-e<sup>ΔK</sup>-βgal and Leu-e<sup>ΔK</sup>-βgal chains would exist as heterodimers within tetramers (Fig. 1B). (This estimate assumes a random assortment of Arg- and Leu-bearing βgal chains in the formation of βgal tetramers. The actual *in vivo* assortment is expected to be biased, to an unknown extent, in favor of homodimeric associations, because individual polysomes would produce βgal chains bearing either Arg or Leu but not both.) In addition, although the e<sup>ΔK</sup> extension (Fig. 2A) is capable of supporting the desired effects, it is also unlikely to be optimal. In summary, the efficacy of this first and necessarily suboptimal bivalent inhibitor bodes well for the future of this design.

A bivalent inhibitor is strikingly more efficacious than an otherwise identical monovalent inhibitor (Figs. 3–5). In addition, our findings are the first evidence that the type 1 and type 2 sites of N-recognin are spatially proximal in the 225-kDa *S. cerevisiae* Ubr1p. While this work was under way, genetic dissection of *S. cerevisiae* Ubr1p identified amino acid residues that are required for the integrity of the type 1 site but not the type 2 site, and *vice versa*.<sup>6</sup> These results provided independent evidence for both the separateness and spatial proximity of the two substrate-binding sites of the 225-kDa N-recognin, in agreement with the present data. Our results (Figs. 3–5) strongly suggest that small bivalent inhibitors of the N-end rule pathway are feasible, and moreover, are expected to be much more potent than their monovalent counterparts. Work to produce such inhibitors is under way.

**Acknowledgments**—We thank members of the Varshavsky laboratory, especially A. Kashina and G. Turner, for helpful discussions and advice in the course of this work. We also thank I. Davydov, F. Du, F. Navarro-Garcia, H. Rao, and especially G. Turner for comments on the manuscript.

#### REFERENCES

- Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12142–12149
- Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234**, 179–186
- Hershko, A. (1991) *Trends Biochem. Sci.* **16**, 265–268
- Peters, J.-M., King, R. W., and Deshaies, R. J. (1998) in *Ubiquitin and the Biology of the Cell*, pp. 345–387, Plenum Press, New York
- Varshavsky, A. (1997) *Trends Biochem. Sci.* **22**, 383–387
- Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30**, 405–439
- Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998) *Cell* **92**, 367–380
- Rechsteiner, M. (1998) in *Ubiquitin and the Biology of the Cell* (Peters, J. M., Harris, J. R., and Finley, D., eds) pp. 147–189, Plenum Press, New York
- Kwon, Y. T., Kashina, A. S., and Varshavsky, A. (1999) *Mol. Cell. Biol.* **19**, 182–193
- Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. G., and Varshavsky, A. (1996) *J. Biol. Chem.* **271**, 28521–28532
- Stewart, A. E., Arfin, S. M., and Bradshaw, R. A. (1995) *J. Biol. Chem.* **270**, 25–28
- Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264**, 16700–16712
- Davydov, I. V., Patra, D., and Varshavsky, A. (1998) *Arch. Biochem. Biophys.* **357**, 317–325
- Bartel, B., Wünning, I., and Varshavsky, A. (1990) *EMBO J.* **9**, 3179–3189
- Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A., and Varshavsky, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7898–7903
- Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2374–2378
- Reiss, Y., Kaim, D., and Hershko, A. (1988) *J. Biol. Chem.* **263**, 2693–2269
- Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17**, 269–277
- Madura, K., and Varshavsky, A. (1994) *Science* **265**, 1454–1458
- Schauber, C., Chen, L., Tongaonkar, P., Vega, I., and Madura, K. (1998) *Genes Cells* **3**, 307–319
- Alagramam, K., Naider, F., and Becker, J. M. (1995) *Mol. Microbiol.* **15**, 225–234
- Ota, I. M., and Varshavsky, A. (1993) *Science* **262**, 566–569
- deGroot, R. J., Rümenapf, T., Kuhn, R. J., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8967–8971
- Sijts, A. J., Pilip, I., and Pamer, E. G. (1997) *J. Biol. Chem.* **272**, 19261–19268
- Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **30**, 280–288
- Taban, C. H., Hondermarck, H., Bradshaw, R. A., and Boilly, B. (1996) *Experientia* **52**, 865–870
- Solomon, V., Lecker, S. H., and Goldberg, A. L. (1998) *J. Biol. Chem.* **273**, 25216–25222
- Solomon, V., Baracos, V., Sarraf, P., and Goldberg, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12602–12607
- Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254**, 1374–1377
- Varshavsky, A., Byrd, C., Davydov, I. V., Dohmen, R. J., Du, F., Ghislain, M., Gonzalez, M., Grigoryev, S., Johnson, E. S., Johnsson, N., Johnston, J. A., Kwon, Y. T., Lévy, F., Lomovskaya, O., Madura, K., Ota, I., Rümenapf, T., Shrader, T. E., Suzuki, T., Turner, G., Waller, P. R. H., and Webster, A. (1998) in *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J. R., and Finley, D., eds) pp. 223–278, Plenum Press, New York
- Kramer, R. H., and Karpen, J. W. (1998) *Nature* **395**, 710–713
- Ptashne, M. (1992) *A Genetic Switch*, Cell Press, Cambridge, MA
- Madura, K., Dohmen, R. J., and Varshavsky, A. (1993) *J. Biol. Chem.* **268**, 12046–12054
- Sherman, F. (1991) *Methods Enzymol.* **194**, 3–21
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (eds) (1996) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York
- Bachmair, A., and Varshavsky, A. (1989) *Cell* **56**, 1019–1032
- Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346**, 287–291
- Sikorski, R. S., and Hieter, P. (1989) *Genetics* **122**, 19–27
- Johnson, E. S., Ma, P. C., Ota, I. M., and Varshavsky, A. (1995) *J. Biol. Chem.* **270**, 17442–17456
- Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) *Nature* **369**, 761–766
- Ghislain, M., Dohmen, R. J., Levy, F., and Varshavsky, A. (1996) *EMBO J.* **15**, 4884–4899
- Lévy, F., Johnsson, N., Rümenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 4907–4912

<sup>6</sup> A. Webster, M. Ghislain, and A. Varshavsky, unpublished data.

# Detection of Transient In Vivo Interactions between Substrate and Transporter during Protein Translocation into the Endoplasmic Reticulum

Martin Dünnwald,\* Alexander Varshavsky,† and Nils Johnsson\*‡

\*Max-Delbrück-Laboratorium, D-50829 Köln, Germany; and †Division of Biology, California Institute of Technology, Pasadena, California 91125

Submitted October 2, 1998; Accepted November 11, 1998

Monitoring Editor: Peter Walter

The split-ubiquitin technique was used to detect transient protein interactions in living cells.  $N_{ub}$ , the N-terminal half of ubiquitin (Ub), was fused to Sec62p, a component of the protein translocation machinery in the endoplasmic reticulum of *Saccharomyces cerevisiae*.  $C_{ub}$ , the C-terminal half of Ub, was fused to the C terminus of a signal sequence. The reconstitution of a quasi-native Ub structure from the two halves of Ub, and the resulting cleavage by Ub-specific proteases at the C terminus of  $C_{ub}$ , serve as a gauge of proximity between the two test proteins linked to  $N_{ub}$  and  $C_{ub}$ . Using this assay, we show that Sec62p is spatially close to the signal sequence of the prepro- $\alpha$ -factor in vivo. This proximity is confined to the nascent polypeptide chain immediately following the signal sequence. In addition, the extent of proximity depends on the nature of the signal sequence.  $C_{ub}$  fusions that bore the signal sequence of invertase resulted in a much lower Ub reconstitution with  $N_{ub}$ -Sec62p than otherwise identical test proteins bearing the signal sequence of prepro- $\alpha$ -factor. An inactive derivative of Sec62p failed to interact with signal sequences in this assay. These in vivo findings are consistent with Sec62p being part of a signal sequence-binding complex.

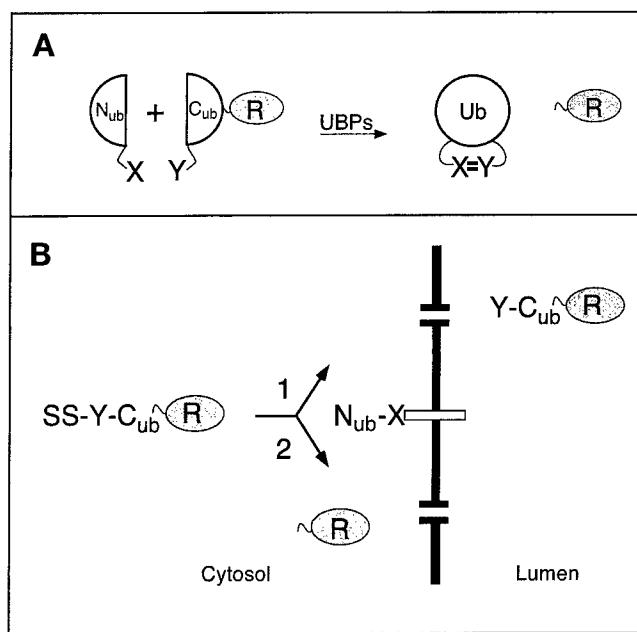
## INTRODUCTION

A critical step during the translocation of a protein across the membrane of the endoplasmic reticulum (ER) is the interaction between the signal sequence of a nascent polypeptide and its receptors (Walter *et al.*, 1981; Gilmore and Blobel, 1985; Walter and Johnson, 1994). A stretch of 8 to 12 hydrophobic residues, often at the N terminus of a protein, comprises a signal sequence that is sufficient to initiate the protein's translocation into the endoplasmic reticulum (ER) (Rapoport *et al.*, 1996). To be compatible with a high flux of polypeptides through a limited number of translocation channels in the ER membrane, the interaction between the signal sequence and its receptors has to be short lived. Its transient nature makes such a receptor-ligand interaction difficult to study, especially in living cells. The approaches used for the

analysis of protein translocation in cell-free systems circumvent the transience of the signal sequence-receptor interaction by pausing or stopping the synthesis of a nascent polypeptide chain at different stages of its movement to and across the ER membrane (Krieg *et al.*, 1986; Kurzchalia *et al.*, 1986; Connolly *et al.*, 1989). Given these constraints, it is essential to verify in vivo the models derived from in vitro studies. The ability to analyze early translocation events in vivo should also be important for defining the immediate environment of the nascent chain on its path from the ribosome to the ER membrane.

Most of the current methods for detecting protein interactions in vivo either do not operate at the ER membrane or are unable to detect a transient proximity between proteins (Fields and Song, 1989; Aronheim *et al.*, 1997; Rossi *et al.*, 1997; Miyawaki *et al.*, 1997). In the present work, we show that the previously developed split-ubiquitin (split-Ub) technique, also called USPS (Ub/split/protein/sensor) (Johnsson and Varshavsky, 1994a, 1997), is capable of detecting a transient in vivo

\* Corresponding author. E-mail address: johnsson@mpiz-koeln.mpg.de.



**Figure 1.** The split-Ub technique and its application to the analysis of protein translocation. (A) N<sub>ub</sub> and C<sub>ub</sub> are linked to the interacting proteins X and Y. The X-Y complex brings N<sub>ub</sub> and C<sub>ub</sub> into close proximity. N<sub>ub</sub> and C<sub>ub</sub> reconstitute a quasi-native Ub moiety, which is cleaved by the UBPs, yielding the free reporter R (Johnsson and Varshavsky, 1994a). (B) Using split-Ub to monitor the proximity between a secretory protein and a component of the translocation machinery. A signal sequence-bearing C<sub>ub</sub> fusion (SS-Y-C<sub>ub</sub>-R) and a N<sub>ub</sub> fusion (N<sub>ub</sub>-X) are coexpressed in a cell. Pathway 1: when N<sub>ub</sub> is linked to a protein not involved in the targeting for translocation, the uncleaved (except for the signal sequence SS) Y-C<sub>ub</sub>-R enters the ER. Pathway 2: when N<sub>ub</sub> is linked to a protein involved in the targeting for translocation the signal sequence of the SS-Y-C<sub>ub</sub>-R brings N<sub>ub</sub> and C<sub>ub</sub> into close contact. As a result, some of the SS-Y-C<sub>ub</sub>-R and N<sub>ub</sub>-X molecules interact to form the quasi-native Ub, yielding the free reporter R in the cytosol.

interaction between polypeptides. The split-Ub method is based on the ability of N<sub>ub</sub> and C<sub>ub</sub>, the N- and C-terminal halves of Ub, to assemble into a quasi-native Ub. Ub-specific proteases (UBPs), which are present in all eukaryotic cells, recognize the reconstituted Ub, but not its halves, and cleave the Ub moiety off a reporter protein that had been linked to the C terminus of C<sub>ub</sub>. The liberation of the reporter serves as a readout indicating the reconstitution of Ub. The assay is designed in a way that prevents efficient association of N<sub>ub</sub> and C<sub>ub</sub> by themselves, but allows it if the two Ub halves are separately linked to proteins that interact in vivo (Figure 1A). The split-Ub assay has been shown to detect the in vivo dimerization of a leucine zipper-containing domain of the Gcn4p transcriptional activator, and the in vivo interaction between two subunits of the oligosaccharyltransferase complex (Johnsson and Varshavsky, 1994a; Stagljar et al., 1998).

In the present work, we focus on the interaction between Sec62p of the yeast *Saccharomyces cerevisiae*

and proteins bearing two different signal sequences. Extensive evidence indicates that Sec62p is a component of the ER translocation machinery (Deshais and Schekman, 1989; Rothblatt et al., 1989; Müsch et al., 1992). Sec62p is a part of the tetrameric Sec62/63 complex that also contains Sec71p and Sec72p (Deshais et al., 1991; Feldheim and Schekman, 1994). Sec62/63p can be isolated as a tetramer, or as a part of a larger assembly, the heptameric Sec complex (Panzner et al., 1995). In addition to the Sec62/63 complex, the heptamer contains the trimer of Sec61p. This trimer (Sec61p, Sss1p, Sbh1p in yeast; Sec61 $\alpha$ , Sec61 $\beta$ , Sec61 $\gamma$  in mammals) forms the aqueous channel through which a polypeptide chain is translocated across the ER membrane (Simon and Blobel, 1991; Görlich et al., 1992; Crowley et al., 1993, 1994; Mothes et al., 1994; Hanein et al., 1996; Beckmann et al., 1997).

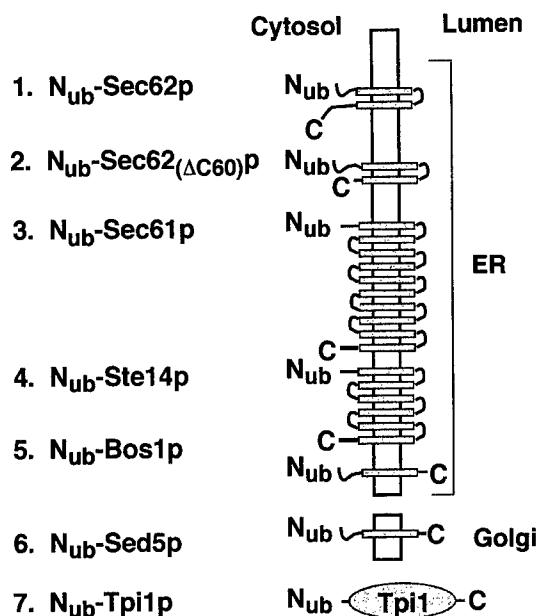
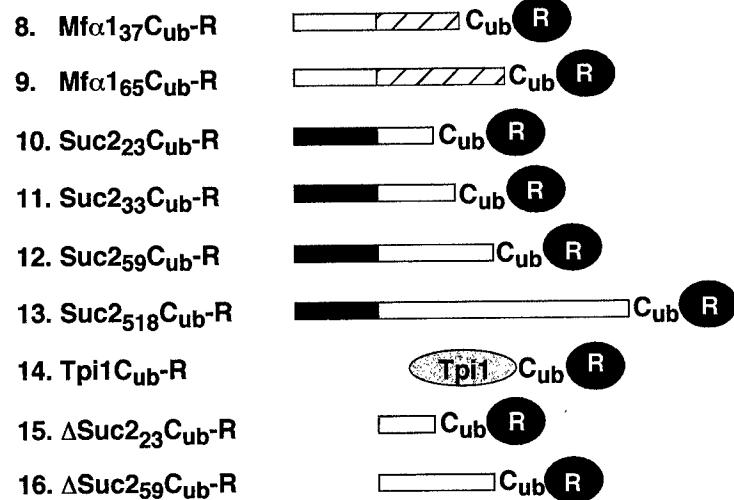
The role of the Sec62/63 tetramer is less well defined. Cross-linking and reconstitution experiments in vitro have shown that Sec62p is close to the nascent polypeptide chain before the initiation of its translocation (Musch et al., 1992; Lyman and Schekman, 1997; Matlack et al., 1997). One important role of Sec63p is its ability to recruit the Hsp70-type protein Kar2p of the ER lumen to the vicinity of a translocating polypeptide (Brodsky and Schekman, 1993; Lyman and Schekman, 1997). The Sec62/63 complex is essential for the post-translational translocation of proteins in reconstituted vesicle preparations (Panzner et al., 1995). Genetic analysis supports this conclusion, by showing that the tetrameric Sec62/63 complex is involved in the translocation of proteins whose targeting to the ER membrane is not abolished by the loss of the signal recognition particle (SRP) (Ng et al., 1996). However, it is less clear whether the Sec62/63 complex is the receptor for the signal sequences of those proteins.

In the present work, we demonstrate the ability of the split-Ub assay to detect transient protein interactions in living cells. We show that the assay can monitor a close proximity between Sec62p and a segment of the nascent chain of a signal sequence-bearing protein. The apparent extent of this proximity is influenced by the nature of the signal sequence and the position of C<sub>ub</sub> in the nascent polypeptide. Our analysis yields a crude map of the environment of the nascent chain during its targeting to and translocation across the ER membrane. Taken together, these findings are the first *in vivo* evidence that Sec62p, a component of the translocation machinery in the endoplasmic reticulum, is a part of a signal sequence-binding complex.

## MATERIALS AND METHODS

### Construction of Test Proteins

The C<sub>ub</sub> fusions 8–13 (Figure 2) were derived from the construct I of Johnsson and Varshavsky (1994a), which encoded Ub-DHFR-ha and

**A N<sub>ub</sub> Fusions****B C<sub>ub</sub> Fusions**

**Figure 2.**  $N_{ub}$  and  $C_{ub}$  test fusions. (A)  $N_{ub}$  (residues 1–36 of Ub) was fused to the N terminus of either a transmembrane protein (constructs 1–6) or a cytosolic protein (construct 7). The N termini of all proteins are located in the cytosol. The orientations and the numbers of membrane-spanning regions (shaded boxes) were derived from the published studies of these proteins, except for Ste14p, for which the exact number of the domains and the localization of the C terminus are not yet known. The  $N_{ub}$  fusions 1–5 are located in the membrane of the ER; the  $N_{ub}$  fusion 6 resides in the membrane of the early Golgi. The  $N_{ub}$  fusion 7 contains the full-length triosephosphate isomerase ( $N_{ub}$ -Tpi1p). (B)  $C_{ub}$  fusions. The  $C_{ub}$  fusions 8 and 9 contain the signal sequence of the prepro- $\alpha$ -factor (shaded boxes), followed by either 37 (construct 8) or 65 residues (construct 9) of the mature  $\alpha$ -factor sequence (striped boxes) and a 7-residue linker sequence (not shown).  $C_{ub}$  fusions 10–13 contain the signal sequence of the Suc2p invertase (dark boxes) followed by 23 (construct 10), 33 (construct 11), 59 (construct 12), or 518 residues (construct 13) of the mature sequence of invertase (open boxes) and a 7-residue linker sequence (not shown). The  $C_{ub}$  fusion 14 contains the complete sequence of *S. cerevisiae* triosephosphate isomerase (Tpi1p) followed by a 17-residue linker peptide and  $C_{ub}$ . The  $C_{ub}$  fusions 15 and 16 are the signal sequence-lacking counterparts of the fusions 10 and 12.  $C_{ub}$  is always followed by a reporter protein. The reporter is DHFR-ha or Ura3p for the  $C_{ub}$  fusions 8–13, and DHFR-ha for the  $C_{ub}$  fusions 14, 15, and 16.

contained a *Bam*HI site at the amino acid position 36 of Ub, and from the previously described Ub translocation constructs I, VI, IX, X, XXIII, and XXV (Johnsson and Varshavsky, 1994b). The above *Bam*HI site of the Ub-DHFR-ha construct I was fused to a linker sequence in which a 5'-*Sall* site allowed the in-frame insertion of an *Eag*I-*Sall* fragment containing the promoter, the signal sequence, and a portion of the mature sequence of the corresponding Ub fusions. The newly introduced sequence was G TCG ACC ATG TCG GGG GGG ATC CCT. The last three triplets encode residues 35, 36, and 37 of Ub (the beginning of  $C_{ub}$ ). The underlined sequences are the *Sall* and *Bam*HI sites, respectively. The final constructs were in the single-copy plasmids pRS314 or pRS315 (Sikorski and Hietter, 1989). Expression of the  $C_{ub}$  fusions bearing Dha (DHFR-ha) as a reporter was mediated by the  $P_{ADH1}$  promoter, except for the  $C_{ub}$  fusion 14, which was expressed from the  $P_{CUP1}$  promoter. The same promoter was used for expressing the Ura3p-based  $C_{ub}$  fusions.

The  $C_{ub}$  fusions 15 and 16 (Figure 2) were derived from constructs 10 and 12 by deleting the *Hind*III fragment spanning the first four codons of the *SUC2* ORF and a short portion of the polylinker sequence between the 3'-end of the  $P_{ADH1}$  promoter and the *SUC2* ORF. As a result, the translation of  $C_{ub}$  fusions 15 and 16 began at the first codon of the mature invertase, skipping its signal sequence. The  $C_{ub}$  fusion 14 (Figure 2) was produced through an in-frame fusion of a PCR fragment containing the complete *TPI1* coding

sequence and  $C_{ub}$ -Dha. The sequence between *TPI1* and  $C_{ub}$  is as follows: AAC GGG TCG ACC GAC TAC AAG GAC GAC GAT GAC AAG GGC TCG ACC ATGTCG GGG GGG ATC CCT. The underlined sequences indicate, respectively, the last codon of *TPI1* and the first three codons of  $C_{ub}$ .

A fragment encoding  $N_{ub}$ -Sec62p was constructed using PCR amplification of a 1050 base pair (bp) fragment containing the *SEC62* ORF. PCR introduced a *Bam*HI site and a linker sequence in front of the start codon of *SEC62* and an *Xho*I site 173 bp downstream of the stop codon. The  $P_{CUP1}$ - $N_{ub}$  modules were cloned as *Bam*HI fragments in frame with the *SEC62* ORF. The sequence between  $N_{ub}$  and *SEC62* is GGG ATC CCT TCT GGG ATG. The first three codons encode residues 35, 36, and 37 of  $N_{ub}$ , followed by the Gly-Ser linker and the start codon of *SEC62*. The *Bam*HI site is underlined. The final constructs resided in pRS316 or pRS313.  $N_{ub}$ -*TPI1*,  $N_{ub}$ -*SED5*,  $N_{ub}$ -*STE14* (a gift from N. Lewke), and  $N_{ub}$ -*Sec62*( $\Delta$ C60)-Dha were constructed similarly to  $N_{ub}$ -*SEC62*. With the exception of  $N_{ub}$ -*Sec62*( $\Delta$ C60)-Dha, which was placed in pRS316 and pRS313, all of these fusions resided in pRS314. The linker connecting codon 35 of  $N_{ub}$  and the first codon of a linked gene was GGG ATC CCT GGG GAT ATG for  $N_{ub}$ -*TPI1* and  $N_{ub}$ -*SED5*, and GGG ATC CCT GGG GAT CAC for  $N_{ub}$ -*STE14*. Underlined are the *Bam*HI site and the first codon of the linked gene. The sequence GGG TCG ACC TTA ATG CAG AGA TCT GGC ATC ATG GTT connected the last codon

**Table 1.** Yeast strains

Strain	Relevant genotype	Source/comment
YPH500	<i>MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52</i>	Sikorski and Hieter (1989)
JD53	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52</i>	Dohmen <i>et al.</i> (1995)
RSY529	<i>MATα his4 leu2-3,112 ura3-52 sec62-1</i>	Deshaines and Schekman (1989)
NJY51	<i>MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52 NUB-BOS1::pRS306</i>	Derivative of YPH500
NJY62-1	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-SEC62::pRS306</i>	Derivative of JD53
NJY73-I	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-BOS1::pRS303</i>	Derivative of JD53
NJY74-I	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-BOS1::pRS306</i>	Derivative of JD53
NJY61-I	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUB-SEC61::pRS304</i>	Derivative of JD53
NJY61-A	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUA-SEC61::pRS304</i>	Derivative of JD53
NJY61-G	<i>MATα his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52 NUG-SEC61::pRS304</i>	Derivative of JD53

of *SEC62* in *N<sub>ub</sub>-Sec62(ΔC60)-Dha* (codon 223, underlined) to the first two codons of DHFR (underlined).

*N<sub>ub</sub>-BOS1* was constructed in part by PCR amplification, with two synthetic oligos and yeast genomic DNA as a template, yielding a 258-bp fragment containing the first 229 bp of the *BOS1* ORF. Upstream of the *BOS1* ATG was a short linker sequence and a *Bam*H site, to allow in-frame fusion of the *P<sub>CUP1</sub>* promoter-*N<sub>ub</sub>* module. The sequence between *N<sub>ub</sub>* and *BOS1* reads: GGG ATC CCT CCA GGA ATG. The first four triplets encode residues 35, 36, 37, and 38 of *N<sub>ub</sub>*, followed by the Gly codon and the start codon of *BOS1*. The *Bam*H site is underlined. The 3'-region of the resulting fragment terminated in a *Sal*I site for insertion into the integrating vectors pRS306 or pRS303. The vector was cut at the unique *Eco*RI site in the *BOS1*-containing fragment and transformed into *S. cerevisiae* strains YPH500 and JD53 to produce, through homologous recombination, the integrated cassette that expressed *N<sub>ub</sub>-Bos1p* from the *P<sub>CUP1</sub>* promoter. The presence of the desired gene fusion and the absence of wild-type *BOS1* were verified by PCR.

An integrated copy of *P<sub>CUP1</sub>-N<sub>ub</sub>-SEC62* was produced by amplifying the first 438 bp of the *SEC62* ORF, and then cloning it, using the *Bam*H and *Eco*RI restriction sites, in frame behind the pRS306-*P<sub>CUP1</sub>-N<sub>ub</sub>* cassette. A unique *Af*III site in the *SEC62* ORF was used to linearize the plasmid for transformation and integration at the *S. cerevisiae SEC62* gene, yielding the strain NJY62-I. The N-terminal 147-residue fragment of Sec62p that was coexpressed with *N<sub>ub</sub>-Sec62p* in the resulting strain has previously been shown to be inactive in translocation (Deshaines and Schekman, 1990). *N<sub>ub</sub>-SEC61* was constructed by targeted integration of a *N<sub>ub</sub>-SEC61*-containing fragment into *SEC61* of the *S. cerevisiae* strain JD53 (Table 1). Specifically, a fragment containing the first 875 bp of the *SEC61* ORF was amplified by PCR and inserted downstream of the pRS304- or pRS303-based *P<sub>CUP1</sub>-N<sub>ub</sub>* cassette, using the flanking *Bam*H and *Eco*RI sites. The linker sequence between *N<sub>ub</sub>* and *SEC61* was GGG ATC CCT GGG TCT GGG ATG. Underlined are the *Bam*H site and the first codon of *SEC61*. For targeted integration, the plasmid was linearized at the unique *Stu*I site in the *SEC61* ORF to create the yeast NJY61-I. A detailed description of the NJY61 strains (Table 1) will be presented elsewhere (Wittke and Johnsson, unpublished data).

All of the *P<sub>CUP1</sub>* promoter-controlled ORFs were expressed under noninducing conditions (no copper added to the medium), except in the experiment shown in Figure 5B, where cells were incubated in the presence of 0.1 mM CuSO<sub>4</sub>.

### Immunoblotting

Proteins fractionated by SDS-12.5% PAGE were electroblotted onto nitrocellulose (Schleicher & Schuell, Dassel, Germany) or polyvinylidene difluoride (Machery-Nagel, Düren, Germany) membranes, using the semidry transfer system (Hoeffer Pharmacia Biotech, San Francisco, CA). Blots were incubated with an anti-ha monoclonal antibody (Babco, Richmond, CA) and visualized using horseradish peroxidase-coupled goat anti-mouse antibody (Bio-Rad, Hercules, CA), the chemiluminescence detection system (Boehringer, Mannheim, Germany), and x-ray films. Where indicated, quantification was performed using the Lumi Imager system (Boehringer).

### Pulse-Chase Analysis

Yeast-rich (YPD) and synthetic minimal media with 2% dextrose (SD) were prepared as described previously (Dohmen *et al.*, 1995). *S. cerevisiae* cells expressing the *N<sub>ub</sub>* and *C<sub>ub</sub>* fusions were grown at 30°C in 10 ml of SD medium without externally added copper to an OD<sub>600</sub> of ~1 and labeled for 5 min with Redivue Promix-[<sup>35</sup>S] (Amersham, Buckinghamshire, United Kingdom), followed (either directly or after a chase) by immunoprecipitation with the anti-ha monoclonal antibody, essentially as described by Johnsson and Varshavsky (1994a,b). The EndoH analysis of glycosylated proteins was carried out as described by Orlean *et al.* (1991). Samples were concentrated before SDS-12.5% PAGE by precipitation with chloroform/methanol. Gels were fixed and enhanced for fluorography. For quantitative analysis, a dried gel was exposed and scanned using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

## RESULTS

### Experimental Strategy

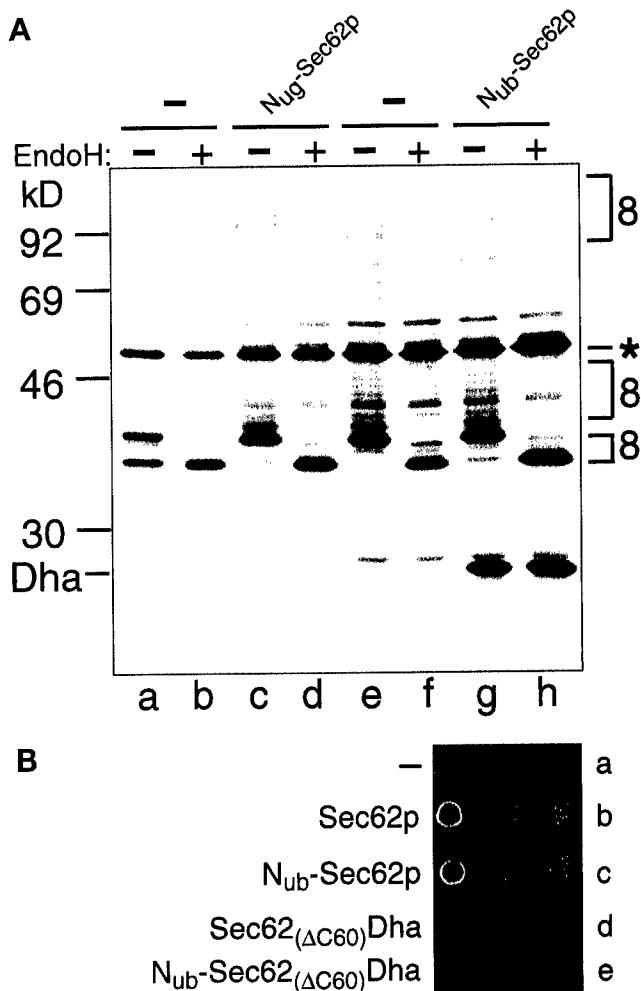
The use of split Ub to monitor the proximity between the proteins X and Y requires the construction of two "complementary" fusion proteins. One fusion bears *N<sub>ub</sub>* (see INTRODUCTION) linked to X (*N<sub>ub</sub>-X*) and the other bears *C<sub>ub</sub>* linked to both Y and a reporter protein R at the C terminus of *C<sub>ub</sub>* (*Y-C<sub>ub</sub>-R*). The liberation of the reporter through the Ub-dependent cleavage by UBP indicates the *in vivo* reconstitution

of a quasi-native Ub from N<sub>ub</sub> and C<sub>ub</sub>. In the split-Ub assay, the efficiency of cleavage at the C terminus of C<sub>ub</sub> in Y-C<sub>ub</sub>-R is measured relative to the efficiency of cleavage observed with selected reference (control) proteins (Figure 1).

To monitor protein interactions during translocation of a protein across the ER membrane, N<sub>ub</sub> was fused to the N terminus of a membrane protein that is a part of the translocation machinery (Figure 1). Owing to the constraint of the assay, which requires the cytosolic location of the reconstituted Ub, the N terminus of this membrane protein must be located in the cell's cytosol. Sec62p has an N-terminal cytosolic domain of 158 residues, which is followed by two membrane-spanning segments and a C-terminal segment also facing the cytosol (Deshaias and Schekman, 1990). N<sub>ub</sub> was therefore fused to the N terminus of Sec62p, yielding N<sub>ub</sub>-Sec62p. C<sub>ub</sub> was sandwiched between the 56 N-terminal residues of the precursor of *S. cerevisiae*  $\alpha$ -factor pheromone (prepro- $\alpha$ -factor) and the ha epitope-tagged mouse dihydrofolate reductase (DHFR-ha; denoted as Dha) as a reporter protein, yielding Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha (Figure 2). The cleavage of the C<sub>ub</sub>-containing fusion at the C<sub>ub</sub>-Dha junction was detected with a monoclonal anti-ha antibody.

#### Split-Ub Detects a Proximity between a Translocating Protein and Sec62p

We first verified that Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha could be translocated across the ER membrane and that the N-terminal extension of Sec62p with N<sub>ub</sub> did not interfere with the Sec62p function in translocation. After a 5-min pulse of wild-type *S. cerevisiae* with  $^{35}$ S-methionine, the labeled Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha was immunoprecipitated as a glycosylated and unclipped fusion (Figure 3A). Thus, Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha could indeed be translocated into the lumen of ER. Introduction of the same Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha construct into the yeast strain RSY529, which carries a temperature-sensitive (ts) variant of Sec62p (Rothblatt *et al.*, 1989), confirmed the severe translocation defect of this strain. About 50% of the pulse-labeled Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha entered the lumen of the ER in this strain at the semipermissive temperature of 30°C, while the rest remained in the cytosol (Figure 3A). Thus, the translocation of Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha depends on Sec62p. This made it possible to determine whether N<sub>ub</sub>-Sec62p is functionally active. The test utilized N<sub>ug</sub>-Sec62p, in which the N-terminal half of Ub contained Gly-13 instead of wild-type Ile-13. This derivative, denoted as N<sub>ug</sub>, has a lower affinity for C<sub>ub</sub> than the wild-type N<sub>ub</sub> (Johnsson and Varshavsky, 1994a). We chose N<sub>ug</sub>-Sec62p for this experiment to minimize the reconstitution of the Ub moiety through interactions between N<sub>ub</sub>-Sec62p and potentially arrested molecules of Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha, which might be localized in the cytosol. Plasmids expressing N<sub>ug</sub>-Sec62p and



**Figure 3.** Sec62p is close to the signal sequence of the  $\alpha$ -factor precursor. (A) *S. cerevisiae* cells expressing Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha (construct 8; Figure 2) were labeled for 5 min with  $^{35}$ S-methionine. The extracted proteins were immunoprecipitated with anti-ha antibody, followed by a mock treatment (lanes a, c, e, and g) or the treatment with EndoH (lanes b, d, f, and h), and SDS-PAGE. The results with cells coexpressing N<sub>ug</sub>- or N<sub>ub</sub>-Sec62p are shown in lanes c and d and g and h, respectively. The analysis was performed with N<sub>ug</sub>-Sec62p in the *S. cerevisiae* mutant RSY529 carrying a ts allele of SEC62 (lanes a-d) or with N<sub>ub</sub>-Sec62p in the wild-type yeast (lanes e-h). Number 8 (following the numbering of the constructs in Figure 2) on the right indicates the positions of uncleaved Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha and its glycosylated forms. An asterisk denotes an unrelated yeast protein that cross-reacts with the anti-ha antibody. (B) N<sub>ub</sub>-Sec62p encodes a functionally active protein. RSY529 cells carrying an empty plasmid (a), Sec62p (b), N<sub>ub</sub>-Sec62p (c), Sec62(ΔC60)-Dha (d), or N<sub>ub</sub>-Sec62(ΔC60)-Dha (e) were spotted on minimal media and grown for 2 d at 30°C (semipermissive temperature for unmodified RSY529).

Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha were cotransformed into RSY529 cells and assayed at 30°C. As in wild-type cells, only translocated Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha, but virtually no free Dha or nontranslocated Mf $\alpha_{37}$ -C<sub>ub</sub>-Dha, was detected after immunoprecipitation and EndoH treatment of the

cells that had been labeled for 5 min with  $^{35}\text{S}$ -methionine (Figure 3A).

To test  $\text{N}_{\text{ub}}$ -Sec62p directly ( $\text{N}_{\text{ub}}$  is the wild-type half of Ub, containing Ile at position 13), we examined its ability to complement the growth defect of RSY529 cells. RSY529 cells expressing  $\text{N}_{\text{ub}}$ -Sec62p were found to grow at the semipermissive temperature of 30°C, in contrast to congenic cells carrying a control plasmid (Figure 3B). To verify that the suppression of the ts phenotype was not due to the initiation of translation from the first (internal) ATG codon of Sec62p within the  $\text{N}_{\text{ub}}$ -Sec62p fusion, the rescue experiment was successfully repeated with the otherwise identical derivative of  $\text{N}_{\text{ub}}$ -Sec62p that lacked the first ATG of SEC62 (our unpublished results).

A significant amount of free Dha was generated when  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha was expressed (in either wild-type or RSY529 cells) together with  $\text{N}_{\text{ub}}$ -Sec62p, which contained the wild-type half of Ub (Figure 3A and our unpublished results). We concluded that Sec62p is close to the nascent polypeptide chain during its translocation into the ER. The cleavage at the C terminus of  $\text{C}_{\text{ub}}$  requires its interaction with  $\text{N}_{\text{ub}}$  and depends on the presence of UBPs (Johnsson and Varshavsky, 1994a). Since UBPs have previously been shown to be absent from the ER (Johnsson and Varshavsky, 1994b), the free Dha moiety had to be produced in the cytosol. Fractionation experiments confirmed that free Dha was absent from membrane-enclosed compartments in whole-cell extracts (our unpublished results). An entirely independent evidence for this conclusion was produced by replacing Dha in  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha with Ura3p as the reporter moiety. Ura3p confers the Ura<sup>+</sup> phenotype on *ura3Δ* cells only if Ura3p has access to the cytosol (Johnsson and Varshavsky, 1994b). In our tests, the cytosolic Ura3p was produced only if  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Ura3p was coexpressed with  $\text{N}_{\text{ub}}$ -Sec62p (compare A and B in Figure 7), in agreement with the other evidence (see above) that the cleavage at the  $\text{C}_{\text{ub}}$ -protein junction takes place exclusively in the cytosol.

The transient nature of the proximity between Sec62p and the nascent chain of a translocated protein was indicated by the near-absence of the released Dha moiety if  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha was coexpressed with either  $\text{N}_{\text{ug}}$ -Sec62p or  $\text{N}_{\text{ua}}$ -Sec62p instead of  $\text{N}_{\text{ub}}$ -Sec62p ( $\text{N}_{\text{ua}}$  denotes Ala at position 13 of  $\text{N}_{\text{ub}}$ ); by contrast, the same experiment with  $\text{N}_{\text{ub}}$ -Sec62p resulted in a significant cleavage of  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha (Figures 3A and 6C). Previous work (Johnsson and Varshavsky, 1994a) has shown that  $\text{N}_{\text{ua}}$  and  $\text{N}_{\text{ug}}$  can induce significant Ub reconstitution when either of them and  $\text{C}_{\text{ub}}$  are linked to polypeptides that form a stable (long-lived) complex in a cell. In summary, the observed absence of significant Ub reconstitution with  $\text{N}_{\text{ua}}$  and  $\text{N}_{\text{ug}}$  (in contrast to  $\text{N}_{\text{ub}}$ ) was interpreted to signify a close but transient (short-lived) proximity between Sec62p and  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha.

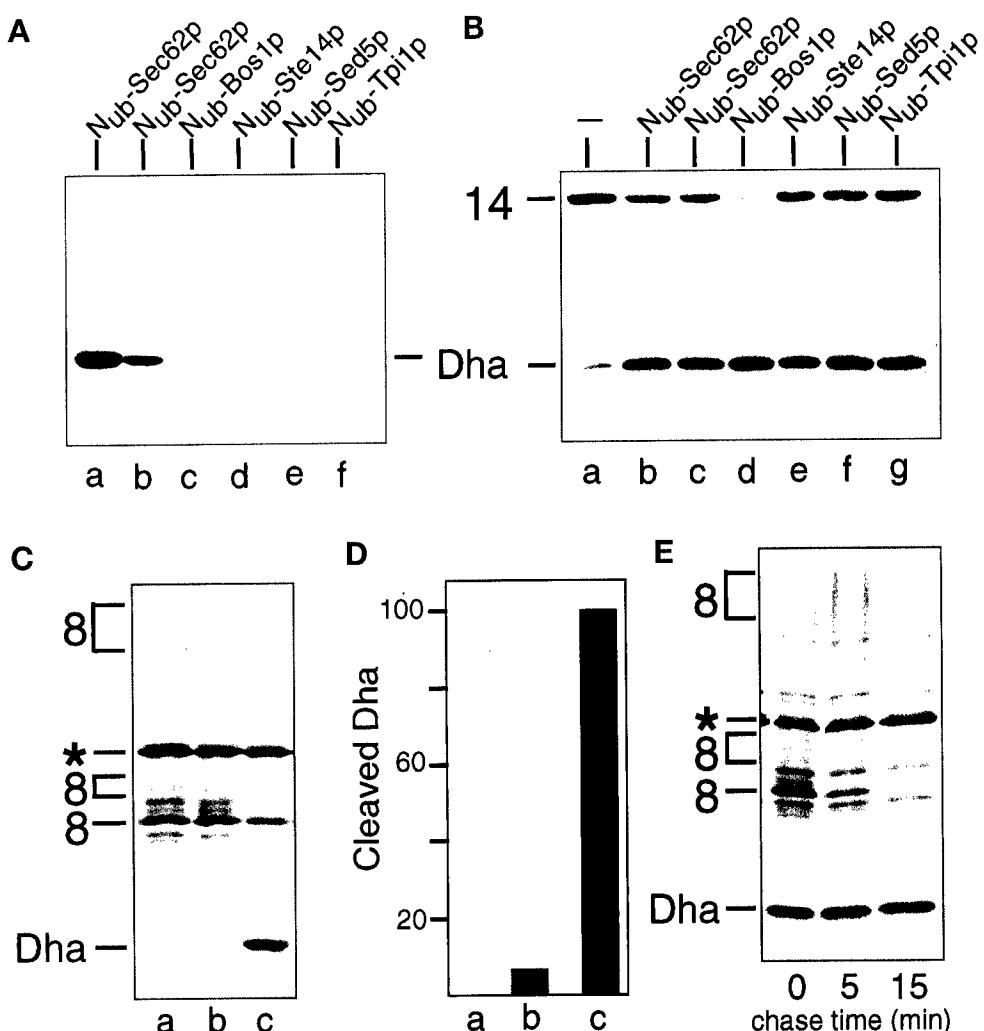
### *Specificity of the Spatial Proximity between a Signal Sequence-bearing Nascent Polypeptide and Sec62p*

A commonly used negative control in a translocation assay is a protein with a defective or absent signal sequence (Allison and Young, 1988; Müsch *et al.*, 1992). Such a control is not entirely compatible with spatio-temporal aspects of the split-Ub assay. Specifically, a  $\text{C}_{\text{ub}}$ -fusion protein lacking a signal sequence accumulates in the cytosol (where the split-Ub assay operates), whereas an analogous signal sequence-bearing protein is continuously removed from this compartment. A direct comparison between reactions that involve a signal sequence-bearing polypeptide and its signal sequence-lacking counterpart requires the ability to compare the local concentrations of the two polypeptides at the site of translocation. We are not aware of an *in vivo* technique that would be independent of the split-Ub assay and at the same time would allow a measurement of these parameters. Therefore, we devised an alternative control. The extent of cleavage of  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha at the  $\text{C}_{\text{ub}}$ -Dha junction should reflect the time-averaged spatial proximity between the nascent  $\text{Mfa}$ -chain and a coexpressed  $\text{N}_{\text{ub}}$ -containing fusion. By comparing the extent of cleavage of  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha in the presence of  $\text{N}_{\text{ub}}$ -Sec62p (Figure 3A) with the analogous activity of  $\text{N}_{\text{ub}}$ -fusion proteins that are not involved in the ER targeting and translocation, we could assess the specificity of the reaction between  $\text{N}_{\text{ub}}$ -Sec62p and  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha.

Four  $\text{N}_{\text{ub}}$ -fusion proteins,  $\text{N}_{\text{ub}}$ -Bos1p,  $\text{N}_{\text{ub}}$ -Ste14p,  $\text{N}_{\text{ub}}$ -Sed5p, and  $\text{N}_{\text{ub}}$ -Tpi1p were tested in the split-Ub assay with  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha. The expected intracellular locations of these  $\text{N}_{\text{ub}}$  fusions, and their predicted topologies in the membrane are shown in Figure 2A (Shim *et al.*, 1991; Banfield *et al.*, 1994; Sapperstein *et al.*, 1994; Lewke and Johnsson, unpublished data). We found that, in contrast to  $\text{N}_{\text{ub}}$ -Sec62p, none of the four tested  $\text{N}_{\text{ub}}$  fusions induced a significant cleavage of  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha (Figure 4A). A small amount of free Dha could be detected in the immunoblots when  $\text{N}_{\text{ub}}$ -Bos1p was overexpressed. The lack of significant Ub reconstitution from  $\text{N}_{\text{ub}}$  and  $\text{C}_{\text{ub}}$  upon coexpression of  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha and the  $\text{N}_{\text{ub}}$ -modified ER membrane proteins, Bos1p, Ste14p (Figure 4A), and Sec12p (Nakanishi *et al.*, 1988; our unpublished results), confirmed that the steady-state concentration of  $\text{Mfa}_{37}\text{-C}_{\text{ub}}$ -Dha in the cytosol was extremely low.

To verify that the observed absence of Ub reconstitution (Figure 4A) was not due to either low concentrations of the tested fusion proteins or reduced accessibility of their linked  $\text{N}_{\text{ub}}$  moieties, we compared the activity of these  $\text{N}_{\text{ub}}$  fusions toward a cytosolic  $\text{C}_{\text{ub}}$ -fusion protein.  $\text{C}_{\text{ub}}$ -Dha was fused to the C terminus of the cytosolic enzyme triosephosphate isomerase

**Figure 4.** The *in vivo* proximity between Sec62p and  $M\alpha_{37}$ -C<sub>ub</sub>-Dha is transient and specific. (A) Immunoblot analysis of extracts of *S. cerevisiae* coexpressing the  $M\alpha_{37}$ -C<sub>ub</sub>-Dha (construct 8; Figure 2) and one of the following constructs: N<sub>ub</sub>-Sec62p, integrated (lane a) or plasmid-borne (lane b); N<sub>ub</sub>-Bos1p (lane c); N<sub>ub</sub>-Ste14p (lane d); N<sub>ub</sub>-Sed5p (lane e); and N<sub>ub</sub>-Tpi1p (lane f). (B) Immunoblot analysis of extracts of *S. cerevisiae* expressing the Tpi1p-C<sub>ub</sub>-Dha fusion (construct 14; Figure 2) alone (lane a) or together with one of the following constructs: N<sub>ub</sub>-Sec62p, either integrated (lane b) or plasmid-borne (lane c); N<sub>ub</sub>-Bos1p (lane d); N<sub>ub</sub>-Ste14p (lane e); N<sub>ub</sub>-Sed5p (lane f); and N<sub>ub</sub>-Tpi1p (lane g). Number 14 on the left indicates the position of uncleaved Tpi1p-C<sub>ub</sub>-Dha. (C) *S. cerevisiae* cells expressing M $\alpha_{37}$ -C<sub>ub</sub>-Dha (construct 8; Figure 2) together with either the vector (lane a), N<sub>ub</sub>-Bos1p (lane b), or N<sub>ub</sub>-Sec62p (lane c) were labeled for 5 min with  $^{35}\text{S}$ -methionine. The extracted proteins were immunoprecipitated with anti-ha antibody and analyzed by SDS-PAGE. (D) Quantitation of the pulse-labeling experiment (C) using Phosphor-Imager. The extent of Dha release in the presence of N<sub>ub</sub>-Sec62p was arbitrarily set at 100. The averages of three experiments are shown. Lanes a, b, and c are the same as in panel C. (E) *S. cerevisiae* cells expressing M $\alpha_{37}$ -C<sub>ub</sub>-Dha together with N<sub>ub</sub>-Sec62p were labeled for 5 min with  $^{35}\text{S}$ -methionine and chased for 5 and 15 min, followed by extraction of proteins, immunoprecipitation with anti-ha antibody, and SDS-PAGE.

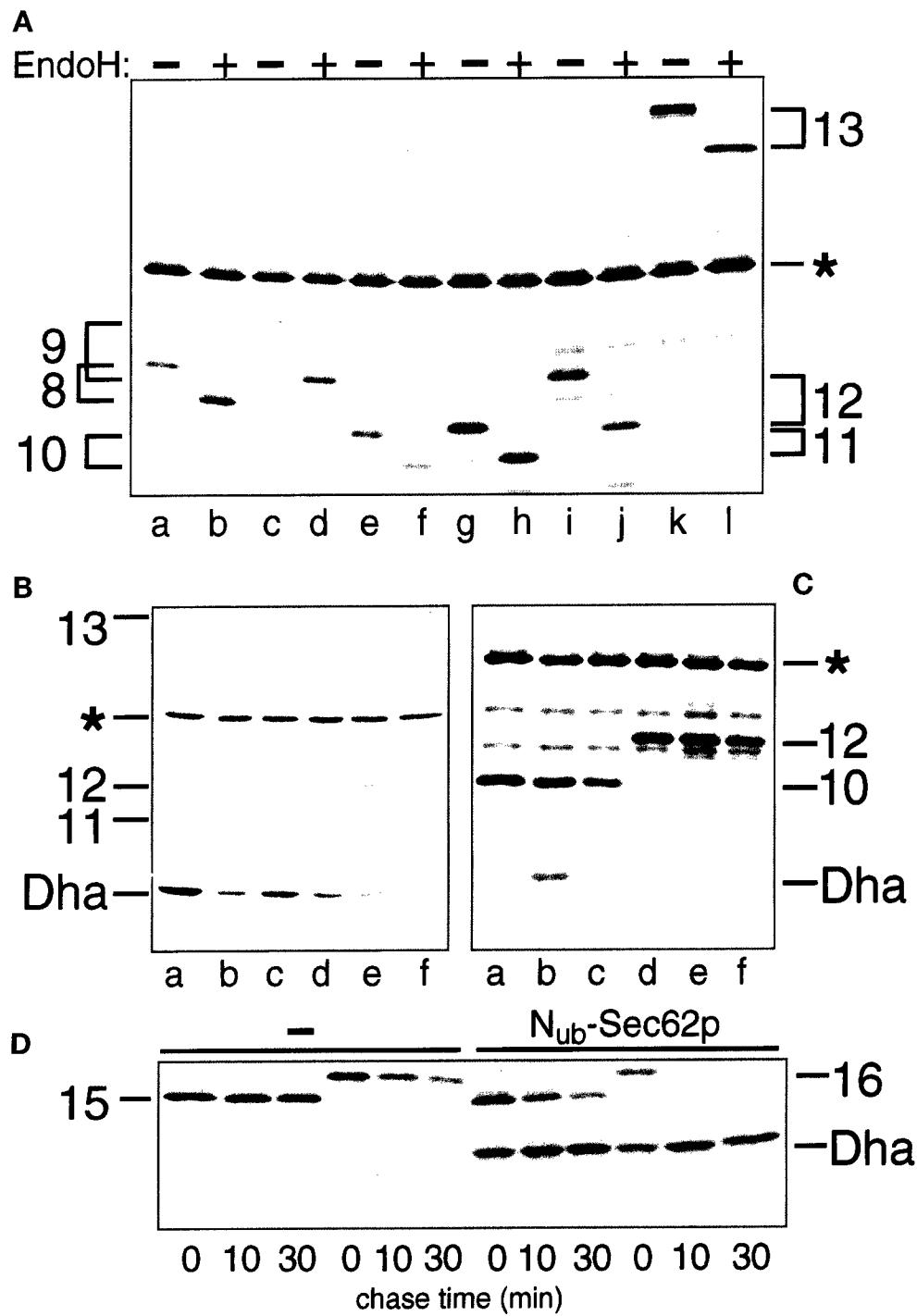


(Tpi1p), yielding Tpi1p-C<sub>ub</sub>-Dha (Figure 2B). All of the N<sub>ub</sub>-fusion proteins in Figure 2A induced a significant release of Dha from the test protein Tpi1p-C<sub>ub</sub>-Dha (Figure 4B). This analysis also suggested that N<sub>ub</sub>-Bos1p was expressed to higher levels than other N<sub>ub</sub> fusions.

To quantify the relative proximities of N<sub>ub</sub>-Bos1p and N<sub>ub</sub>-Sec62p to M $\alpha_{37}$ -C<sub>ub</sub>-Dha, yeast cells were labeled for 5 min with  $^{35}\text{S}$ -methionine, and the released Dha was determined as described in the legend to Figure 4. Coexpression of N<sub>ub</sub>-Sec62p and M $\alpha_{37}$ -C<sub>ub</sub>-Dha yielded ~15 times more of the free Dha than coexpression of N<sub>ub</sub>-Bos1p and M $\alpha_{37}$ -C<sub>ub</sub>-Dha (Figure 4, C and D). Assuming that the N<sub>ub</sub> moieties in N<sub>ub</sub>-Sec62p and N<sub>ub</sub>-Bos1p were equally accessible to the cytosol (Figure 4B), we concluded that the time-averaged proximity between the nascent chain of

M $\alpha_{37}$ -C<sub>ub</sub>-Dha and the N<sub>ub</sub>-bearing transmembrane proteins was much higher for Sec62p than for the ER membrane proteins that are not involved in targeting or translocation. Note that this analysis may actually underestimate the proximity of Sec62p to the nascent chain, because we invariably observed a more efficient cleavage of M $\alpha_{37}$ -C<sub>ub</sub>-Dha when N<sub>ub</sub>-Sec62p was the only form of Sec62p in the cell (Figures 4A and 6A). Therefore we interpret the reduced cleavage of M $\alpha_{37}$ -C<sub>ub</sub>-Dha in the presence of both N<sub>ub</sub>-Sec62p and the native Sec62p as the consequence of competition between those two Sec62p-containing species for either the signal sequences of translocated proteins or the ligands of Sec62p in the complex of Sec proteins.

Recent evidence indicates that misfolded or otherwise abnormal proteins in the lumen of the ER can be retrotransported across the ER membrane back into



**Figure 5.** The nature of the signal sequence and its distance from  $C_{ub}$  determine the extent of cleavage of  $C_{ub}$ -R in the presence of  $N_{ub}$ -Sec62p. (A) *S. cerevisiae* expressing  $Mf\alpha_{37}C_{ub}$ -Dha (construct 8; Figure 2) (lanes a and b),  $Mf\alpha_{65}C_{ub}$ -Dha (construct 9) (lanes c and d),  $Suc2_{23}C_{ub}$ -Dha (construct 10) (lanes e and f),  $Suc2_{33}C_{ub}$ -Dha (construct 11) (lanes g and h),  $Suc2_{59}C_{ub}$ -Dha (construct 12) (lanes i and j) and  $Suc2_{518}C_{ub}$ -Dha (construct 13) (lanes k and l) were labeled with  $^{35}\text{S}$ -methionine for 5 min. The extracted proteins were either mock-treated (lanes a, c, e, g, i, and k) or treated with EndoH (lanes b, d, f, h, j, and l), followed by immunoprecipitation with anti-ha antibody and SDS-PAGE. (B) Same as panel A but the cells also contained  $N_{ub}$ -Sec62p in addition to the Cub-fusions  $Mf\alpha_{37}C_{ub}$ -Dha,  $Mf\alpha_{65}C_{ub}$ -Dha,  $Suc2_{23}C_{ub}$ -Dha,  $Suc2_{33}C_{ub}$ -Dha,  $Suc2_{59}C_{ub}$ -Dha,  $Suc2_{518}C_{ub}$ -Dha (lanes a-f). The analysis was carried out by immunoblotting whole-cell extracts with the anti-ha antibody. (C) *S. cerevisiae* cells expressing  $Suc2_{23}C_{ub}$ -Dha (construct 10; Figure 2) (lanes a-c) and  $Suc2_{59}C_{ub}$ -Dha (construct 12; Figure 2) (lanes d-f) together with either  $N_{ub}$ -Sec62p (lanes b and e),  $N_{ub}$ -Bos1p (lanes c and f) or the vector (lanes a and d) were labeled for 5 min with  $^{35}\text{S}$ -methionine. Whole-cell extracts were immunoprecipitated with anti-ha antibody, followed by SDS-PAGE and autoradiography. (D) *S. cerevisiae* cells expressing  $\Delta Suc2_{23}C_{ub}$ -Dha (construct 15; Figure 2) or  $\Delta Suc2_{59}C_{ub}$ -Dha (construct 16; Figure 2) together with either the vector (first six lanes) or  $N_{ub}$ -Sec62p (last six lanes) were labeled for 5 min with  $^{35}\text{S}$ -methionine and chased for 10 and 30 min, followed by extraction, immunoprecipitation with anti-ha antibody, and SDS-PAGE. Numbers 15 and 16 indicate the positions of the corresponding (uncleaved)  $C_{ub}$  fusions.

the cytosol, where they are degraded by the Ub system (Biederer *et al.*, 1996; Hiller *et al.*, 1996; Wiertz *et al.*, 1996). This retrotransport involves at least some of the components of the known ER translocation machinery (Plemper *et al.*, 1997). To determine whether the cleavage of  $Mf\alpha_{37}C_{ub}$ -Dha at the  $C_{ub}$ -Dha junction occurs

during translocation into the ER or during (in this case) a hypothetical retrotransport from the ER, cells coexpressing  $N_{ub}$ -Sec62p and  $Mf\alpha_{37}C_{ub}$ -Dha were labeled for 5 min with  $^{35}\text{S}$ -methionine, and then chased for 15 min (Figure 4E). Although the translocated  $Mf\alpha_{37}C_{ub}$ -Dha disappeared rapidly during the chase,

the amount of free Dha that accumulated during the pulse remained constant.

We conclude that the *in vivo* proximity between Sec62p and Mfa<sub>37</sub>-C<sub>ub</sub>-Dha that is detected by the split-Ub assay occurs either during or very shortly after the synthesis of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha. The apparent disappearance of the pulse-labeled, translocated Mfa<sub>37</sub>-C<sub>ub</sub>-Dha during the chase accounts for the difficulty in detecting this species by a steady-state assay such as immunoblotting (Figures 4A, 5B, and 6A). The likely cause of the disappearance of translocated Mfa<sub>37</sub>-C<sub>ub</sub>-Dha is its molecular mass heterogeneity, owing to its glycosylation, which results in a smear upon SDS-PAGE (Figures 3A, 4C, and 4E).

#### *The Efficiency of Ub Reconstitution Mediated by N<sub>ub</sub>-Sec62p Depends on Both the Identity of a Signal Sequence and the Position of C<sub>ub</sub> in the Nascent Polypeptide Chain*

The proximity of Sec62p to the signal sequence of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha is detected, in the split-Ub assay, through the ability of N<sub>ub</sub>-Sec62p to induce the cleavage of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha at the C<sub>ub</sub>-Dha junction (Figure 3A). If this cleavage reflects the physical proximity between Sec62p and a signal sequence, the efficiency of cleavage should decrease if the C<sub>ub</sub> moiety is moved closer to the C terminus of the nascent polypeptide chain. However, this purely spatial consideration neglects the temporal aspect of the translocation process (Walter and Johnson, 1994). The targeting and the actual translocation are initiated during or shortly after the synthesis of a signal sequence-bearing protein. Consequently, the C-terminal parts of the nascent chain may still be synthesized, or at least associated with the ribosome, at the time when Sec62p and the signal sequence have already become spatially close. Extending the spacer would increase the distance between C<sub>ub</sub> and the signal sequence of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha. This would be expected to decrease the time window available for the interaction between the C<sub>ub</sub> moiety of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha and the N<sub>ub</sub> moiety of N<sub>ub</sub>-Sec62p. Therefore, a test of this kind cannot deconvolute the contribution of each of the two parameters (increased spatial distance along the chain between Sec62p and C<sub>ub</sub> and decreased time window for the N<sub>ub</sub>-C<sub>ub</sub> interaction) to the overall effect of extending the length of the polypeptide between the signal sequence and C<sub>ub</sub>. These constraints notwithstanding, moving the C<sub>ub</sub> moiety of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha further away from its signal sequence makes it possible to gauge the accessibility of Sec62p to specific regions of the nascent polypeptide chain *in vivo*.

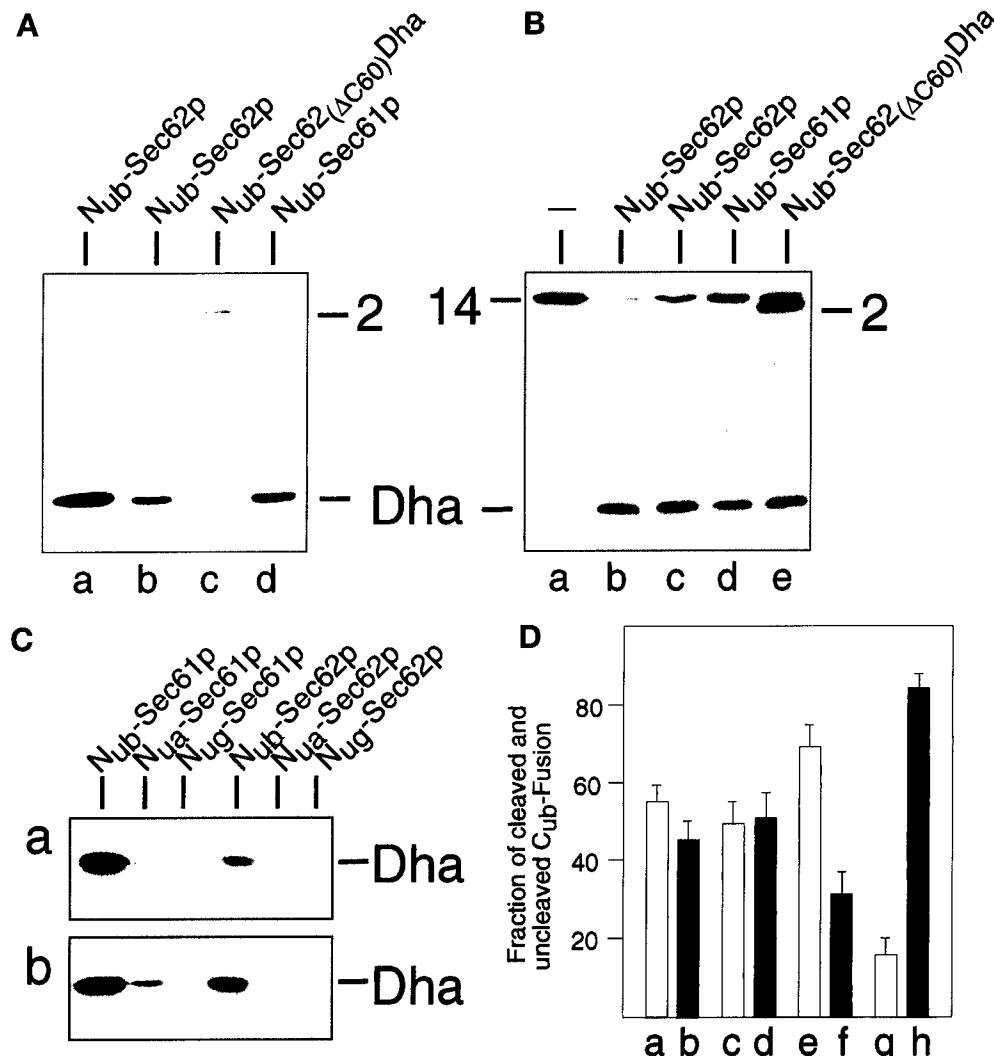
In the actual experiment, the distance between the signal sequence of Mfa<sub>37</sub>-C<sub>ub</sub>-Dha and its C<sub>ub</sub> moiety was increased from 37 to 65 residues (Mfa<sub>65</sub>-C<sub>ub</sub>-Dha; Figure 2B, construct 9). The results of EndoH treat-

ment of Mfa<sub>65</sub>-C<sub>ub</sub>-Dha immunoprecipitated from pulse-labeled wild-type cells confirmed that Mfa<sub>65</sub>-C<sub>ub</sub>-Dha was efficiently translocated into the ER (Figure 5A). However, the efficiency of the Dha-yielding cleavage of Mfa<sub>65</sub>-C<sub>ub</sub>-Dha upon coexpression of N<sub>ub</sub>-Sec62p was clearly reduced in comparison to the same cleavage with Mfa<sub>37</sub>-C<sub>ub</sub>-Dha and N<sub>ub</sub>-Sec62p (Figure 5B).

Both the kinetics and the mode of targeting for translocation are influenced by the identity of a signal sequence (Bird *et al.*, 1987; Johnsson and Varshavsky, 1994b; Ng *et al.*, 1996). For example, the efficient translocation of invertase (Suc2p) requires the SRP, in contrast to a much weaker requirement for SRP in the case of the prepro- $\alpha$ -factor's signal sequence (Hann and Walter, 1991; Ogg *et al.*, 1992; Johnsson and Varshavsky, 1994b). Consequently, the coupling between translation and translocation is tighter for proteins bearing the invertase signal sequence than for proteins carrying the signal sequence of the  $\alpha$ -factor.

We assessed the *in vivo* proximity of the invertase signal sequence to Sec62p by measuring the reconstitution of Ub from N<sub>ub</sub>-Sec62p and Suc2-C<sub>ub</sub>-Dha, where the Suc2p moiety was linked to C<sub>ub</sub> through a spacer of increasing length (Figure 2B). The expression and efficient translocation of different Suc2-C<sub>ub</sub>-Dha constructs were assayed by immunoprecipitation and subsequent EndoH treatment (Figure 5A). The proximity of C<sub>ub</sub> in Suc2-C<sub>ub</sub>-Dha to N<sub>ub</sub> of N<sub>ub</sub>-Sec62p was assayed by immunoblot detection of the cleavage-derived free Dha in whole-cell extracts. The pattern already observed for the Mfa-C<sub>ub</sub>-Dha constructs recurred with the constructs bearing the invertase signal sequence (Figure 5B). Moreover, coexpression of N<sub>ub</sub>-Sec62p with either Suc2<sub>23</sub>-C<sub>ub</sub>-Dha or Suc2<sub>33</sub>-C<sub>ub</sub>-Dha yielded lower amounts of free Dha than the analogous assays with N<sub>ub</sub>-Sec62p and Mfa<sub>37</sub>-C<sub>ub</sub>-Dha, which bears a spacer of comparable length (Figure 5B).

Pulse-chase analyses with cells expressing N<sub>ub</sub>-Sec62p (or N<sub>ub</sub>-Bos1p) and either Suc2<sub>23</sub>-C<sub>ub</sub>-Dha or Suc2<sub>59</sub>-C<sub>ub</sub>-Dha confirmed the immunoblot data. Specifically, a significant release of free Dha was observed only for the pair of N<sub>ub</sub>-Sec62p and Suc2<sub>23</sub>-C<sub>ub</sub>-Dha (Figure 5C). Our previous work has shown that the segment of the nascent polypeptide chain where the C<sub>ub</sub> moiety was inserted in either the Suc2<sub>23</sub>-C<sub>ub</sub>-Dha or the Mfa<sub>37</sub>-C<sub>ub</sub>-Dha fusion is transiently exposed to the cytosol—until the initiation of ER translocation (Johnsson and Varshavsky, 1994b). Therefore, we compared the ratios of cleaved to uncleaved Suc2<sub>23</sub>-C<sub>ub</sub>-Dha and Mfa<sub>37</sub>-C<sub>ub</sub>-Dha. Cells expressing N<sub>ub</sub>-Sec62p and either the C<sub>ub</sub> fusion 8 or 10 (Figure 2B) were labeled for 5 min with <sup>35</sup>S-methionine and processed for immunoprecipitation with anti-ha antibody, followed by determination of the cleaved-to-uncleaved ratio (Figure 6D). This ratio, a measure of the time-averaged proximity of Sec62p to a translocating protein,



(black bars; b, d, f, and h)  $\text{C}_{\text{ub}}$  fusions. The sum of a cleaved and uncleaved fusion was set at 100 in each of the three independent experiments. SDs are also indicated.

was ~eightfold higher for a nascent polypeptide bearing the signal sequence of  $\alpha$ -factor than for a nascent polypeptide bearing the invertase signal sequence (Figure 6D; compare Figures 4C and 5C).

Spacer sequences of different length or composition upstream of the  $\text{C}_{\text{ub}}$  moiety might nonspecifically influence the interaction between  $\text{N}_{\text{ub}}$  and  $\text{C}_{\text{ub}}$ . To assess this potential spacer effect, we constructed signal sequence-lacking versions of  $\text{Suc}2_{23}\text{-C}_{\text{ub}}$ -Dha and  $\text{Suc}2_{59}\text{-C}_{\text{ub}}$ -Dha (Figure 2B,  $\text{C}_{\text{ub}}$  fusions 15 and 16), and compared their ability to reconstitute Ub in the presence of coexpressed  $\text{N}_{\text{ub}}\text{-Sec}62\text{p}$ . Both of these  $\text{C}_{\text{ub}}$  fusions were cleaved at the  $\text{C}_{\text{ub}}$ -Dha junction at approximately the same rate in the presence of  $\text{N}_{\text{ub}}\text{-Sec}62\text{p}$  (Figure 5D), in contrast to the marked difference in the rate of cleavage observed for their signal sequence-bearing counterparts (Figure 5, B and C). This control exper-

iment further emphasized the effect of distance between a signal sequence and the  $\text{C}_{\text{ub}}$  moiety on the efficiency of Ub reconstitution in the presence of  $\text{N}_{\text{ub}}\text{-Sec}62\text{p}$ . We conclude that the accessibility of Sec62p *in vivo* to a specific region of the nascent polypeptide chain is influenced by both the nature of a signal sequence and its distance from that region.

#### *Sec61p Is Equidistant from Two Different Signal Sequences*

A direct comparison between two different signal sequences upstream of the  $\text{C}_{\text{ub}}$  moiety presumes approximately equal residence times of the corresponding  $\text{C}_{\text{ub}}$  moieties in the cytosol. It is also essential to know that the influence of the identity of a signal sequence on the rate of Ub reconstitution is not due to

a nonspecific intramolecular interaction. One way to address these issues involves measuring the reconstitution of Ub from the C<sub>ub</sub> moieties of the fusions 8 and 10 (Figure 2B) and a N<sub>ub</sub>-containing fusion that is not involved in translocation. As illustrated in Figure 4, this test is not feasible because of the rapid translocation of C<sub>ub</sub>-containing constructs into the ER. Note, however, that since Sec61p is the central component of the translocation pore, proteins that utilize different targeting pathways will converge at Sec61p shortly before their translocation (Jungnickel and Rapoport, 1995). Taking advantage of this property of Sec61p, we assayed the proximity of Mfα<sub>37</sub>-C<sub>ub</sub>-Dha and Suc2<sub>23</sub>-C<sub>ub</sub>-Dha to N<sub>ub</sub>-Sec61p. If Mfα<sub>37</sub>-C<sub>ub</sub>-Dha and Suc2<sub>23</sub>-C<sub>ub</sub>-Dha are cleaved at the C<sub>ub</sub>-Dha junction equally well in the presence of N<sub>ub</sub>-Sec61p, the above interpretation of the observed selectivity of N<sub>ub</sub>-Sec62p toward Mfα<sub>37</sub>-C<sub>ub</sub>-Dha (Figures 5 and 6) would be confirmed.

To carry out this test, N<sub>ub</sub> was fused to the cytosolic N terminus of Sec61p (Figure 2A) (Wilkinson *et al.*, 1996). N<sub>ub</sub>-Sec61p is functionally active (Wittke and Johnsson, unpublished data). It induced the release of free Dha from of Mfα<sub>37</sub>-C<sub>ub</sub>-Dha and Tpi1p-Cub-Dha with efficiency similar to that of N<sub>ub</sub>-Sec62p (Figure 6, A and B). Thus, the split-Ub assay independently confirmed that Sec61p is close to the nascent polypeptide chain during its translocation. To compare the *in vivo* interactions of Sec61p with the C<sub>ub</sub> fusions 8 and 10, which bore different signal sequences (Figure 2B), the amount of free Dha was determined by immunoblotting of whole-cell extracts. It was found that in the presence of N<sub>ub</sub>-Sec61p, similar amounts of Dha were released from Mfα<sub>37</sub>-C<sub>ub</sub>-Dha and Suc2<sub>23</sub>-C<sub>ub</sub>-Dha, whereas in the presence of N<sub>ub</sub>-Sec62p twice as much Dha was released from Mfα<sub>37</sub>-C<sub>ub</sub>-Dha than from Suc2<sub>23</sub>-C<sub>ub</sub>-Dha (Figure 6C).

This result was confirmed and extended by labeling the cotransformed cells for 5 min with <sup>35</sup>S-methionine and quantifying the ratio of cleaved-to-uncleaved C<sub>ub</sub> fusions (Figure 6D). As was already observed by the immunoblot analysis, the above ratio was ~1 for both Mfα<sub>37</sub>-C<sub>ub</sub>-Dha and Suc2<sub>23</sub>-C<sub>ub</sub>-Dha in the presence of N<sub>ub</sub>-Sec61p, but ~2 for Mfα<sub>37</sub>-C<sub>ub</sub>-Dha, and ~0.25 for Suc2<sub>23</sub>-C<sub>ub</sub>-Dha in the presence of N<sub>ub</sub>-Sec62p (Figure 6D). The difference revealed by the pulse-immunoprecipitation analysis is higher than the estimate obtained by the immunoblot analysis, most likely because of the continuous accumulation of cleaved (and long-lived) Dha before the processing of cells for immunoblotting.

#### *A C-terminally Truncated Sec62p Is No Longer Proximal to the Signal Sequence*

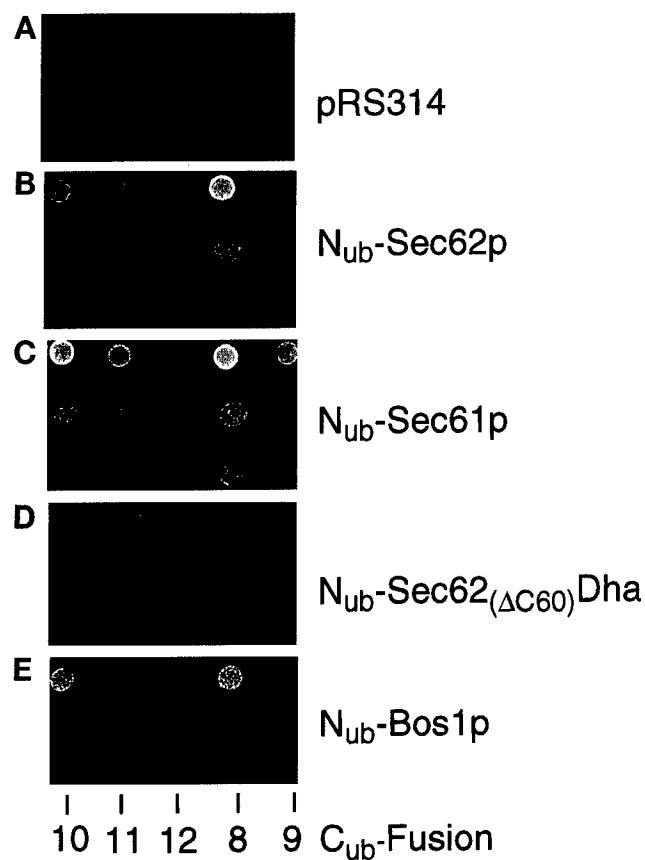
Does the proximity of Sec62p to a nascent polypeptide chain that is detected by the split-Ub assay reflect the

physical binding of the signal sequence to this protein? We constructed a derivative of N<sub>ub</sub>-Sec62p in which the C-terminal 60 residues of Sec62p were replaced by the DHFR-ha (Dha) moiety, yielding Sec62(ΔC60)-Dha. A similar Sec62p-invertase fusion was described by Deshaies and Schekman (1990) and shown to be nonfunctional. As expected, neither Sec62(ΔC60)-Dha nor N<sub>ub</sub>-Sec62(ΔC60)-Dha complemented the ts phenotype of RSY529 cells (Figure 3B).

The Ub-reconstitution activity of N<sub>ub</sub>-Sec62(ΔC60)-Dha in the presence of either Mfα<sub>37</sub>-C<sub>ub</sub>-Dha or Tpi1-C<sub>ub</sub>-Dha (Figure 6, A and B) was compared with the activity of N<sub>ub</sub>-Sec62p and N<sub>ub</sub>-Sec61p in the presence of the same C<sub>ub</sub>-containing fusions. Remarkably, no cleavage of Mfα<sub>37</sub>-C<sub>ub</sub>-Dha was observed in the presence of N<sub>ub</sub>-Sec62(ΔC60)-Dha, whereas the cytosolic Tpi1-C<sub>ub</sub>-Dha was cleaved. This result (Figure 6, A and B) indicated that the concentration and accessibility of N<sub>ub</sub> were comparable for the functionally inactive N<sub>ub</sub>-Sec62(ΔC60)-Dha and the functionally active N<sub>ub</sub>-Sec62p. In these experiments, N<sub>ub</sub>-Sec62(ΔC60)-Dha, which could be detected with the anti-ha antibody (Figure 6, A and B), was expressed from the uninduced P<sub>CUP1</sub> promoter. Strikingly, even overexpression of N<sub>ub</sub>-Sec62(ΔC60)-Dha, from the copper-induced P<sub>CUP1</sub>, did not result in a significant cleavage of Mfα<sub>37</sub>-C<sub>ub</sub>-Dha (our unpublished results). These control experiments with the inactive derivative of Sec62p indicated that the proximity signal in the split-Ub assay with Sec62p requires the functional activity of Sec62p.

#### *Using Ura3p Reporter to Detect the In Vivo Proximity between Sec62p and Signal Sequences*

The DHFR-ha (Dha) reporter moiety of Mfα<sub>37</sub>-C<sub>ub</sub>-Dha was replaced by *S. cerevisiae* Ura3p (orotidine-5'-phosphate decarboxylase), yielding Mfα<sub>37</sub>-C<sub>ub</sub>-Ura3p. The use of cytosolic Ura3p as a reporter for translocation across membranes is well documented (Maarse *et al.*, 1992; Johnsson and Varshavsky, 1994b; Ng *et al.*, 1996). The high sensitivity of Ura3p-based assays (cells become Ura<sup>+</sup> if a threshold amount of Ura3p is present in the cytosol) allowed us to express the N<sub>ub</sub> and C<sub>ub</sub> fusions from the uninduced P<sub>CUP1</sub> promoter. Since the efficient translocation of Mfα<sub>37</sub>-C<sub>ub</sub>-Ura3p sequesters the Ura3p activity in the ER, a *ura3Δ* strain of *S. cerevisiae* that expressed Mfα<sub>37</sub>-C<sub>ub</sub>-Ura3p remained Ura<sup>-</sup> (Figure 7A). N<sub>ub</sub>-Sec62p, which, as shown above, is close to the nascent chain of Mfα<sub>37</sub>-C<sub>ub</sub>-Dha during its translocation, induced enough cleavage of Mfα<sub>37</sub>-C<sub>ub</sub>-Ura3p at the C<sub>ub</sub>-Ura3p junction to render cells Ura<sup>+</sup> (Figure 7B). Cells were transformed with either N<sub>ub</sub>-Sec62p, N<sub>ub</sub>-Sec61p, N<sub>ub</sub>-Sec62(ΔC60)-Dha, or N<sub>ub</sub>-Bos1p to compare relative proximities of these N<sub>ub</sub>-containing proteins to C<sub>ub</sub> fusions bearing the Ura3p reporter moiety and either



**Figure 7.** The use of a metabolic marker to assess the proximity between a component of the translocation machinery and a translocated protein. *S. cerevisiae* expressing the *C<sub>ub</sub>* fusions 8–12 (Figure 2) that contained Ura3p instead of Dha (see the main text) were transformed with the vector (A) or plasmids expressing N<sub>ub</sub>-Sec62p (B), N<sub>ub</sub>-Sec61p (C), N<sub>ub</sub>-Sec62(ΔC60)Dha (D), and N<sub>ub</sub>-Bos1p (E). Cells were grown in a liquid uracil-containing SD medium, and ~10<sup>5</sup>, 10<sup>3</sup>, and 10<sup>2</sup> cells were spotted onto uracil-lacking SD medium. Plates were examined after 18 h at 30°C.

the invertase-derived or the  $\alpha$ -factor-derived signal sequence (Figure 7, B–E). The cells were spotted on plates lacking uracil and incubated at 30°C for 18 h. The growth patterns of strains that expressed different combinations of N<sub>ub</sub>- and C<sub>ub</sub>-containing fusions confirmed the results of analyses with analogous (but more highly expressed) Dha-based constructs.

In particular, the interaction of Sec62p with the signal sequence of prepro- $\alpha$ -factor was stronger than with the signal sequence of invertase. This proximity was not detectable when the distance between a signal sequence and the C<sub>ub</sub> moiety of a fusion was increased (Figure 7B). Sec61p appears to be equally close to both of the signal sequences tested. Again, the proximity signal was gradually lost when the distance between the signal sequence and the C<sub>ub</sub> moiety was increased (Figure 7C). Cells acquired a weak Ura<sup>+</sup> phenotype in the presence of N<sub>ub</sub>-Bos1p and the C<sub>ub</sub>-Ura3p fusions

8 and 10 (Figure 7E). If used as a reference to discriminate between specific and nonspecific signals in this assay, Sec62p, under these conditions, appears to interact only with the  $\alpha$ -factor signal sequence. No interaction with any of the tested C<sub>ub</sub> constructs was detectable with the functionally inactive N<sub>ub</sub>-Sec62(ΔC60)-Dha (Figure 7D).

## DISCUSSION

The new application of the split-Ub technique (Johnson and Varshavsky, 1994a, 1997) described in the present work introduces a tool for the analysis of transient (short-lived) protein interactions in living cells. A split-Ub assay involves the tagging of two (presumably) interacting proteins with the N- and C-terminal halves of Ub, N<sub>ub</sub> and C<sub>ub</sub>, and monitoring, in a variety of ways, the release of the reporter protein fused to the C terminus of C<sub>ub</sub>. The reporter release, through the cleavage by Ub-specific processing proteases (UBPs), takes place in the cytosol if the two halves of Ub interact in vivo to form a quasi-native Ub moiety upstream of the reporter (Figure 1). Among the advantages of this method are its applicability either in living cells or in vitro and its sensitivity to kinetic aspects of a protein interaction.

In the present work, we applied the split-Ub technique to the problem of protein translocation across membranes. We showed that Sec62p of *S. cerevisiae* is spatially close to the signal sequence of the nascent  $\alpha$ -factor polypeptide in vivo. This proximity is confined to the nascent polypeptide chain immediately following the signal sequence. In addition, the extent of proximity depends on the nature of the signal sequence. Specifically, C<sub>ub</sub>-containing test proteins that bore the signal sequence of invertase resulted in a much lower Ub reconstitution with N<sub>ub</sub>-Sec62p than the same C<sub>ub</sub>-containing proteins bearing the signal sequence of  $\alpha$ -factor. An inactive derivative of Sec62p failed to interact with signal sequences in the split-Ub assay. Taken together, these findings are the first in vivo evidence that *S. cerevisiae* Sec62p, a component of the ER translocation machinery, is a part of a signal sequence-binding complex.

### *In Vivo Proximity between Sec62p and the Signal Sequence of $\alpha$ -Factor*

We have previously shown that a region of the nascent polypeptide chain that lies close to the signal sequence of invertase or the prepro- $\alpha$ -factor is briefly exposed to the cytosol before its translocation into the ER (Johnson and Varshavsky, 1994b). This feature of translocation enabled us, in the present work, to use the split-Ub assay for monitoring the proximity between a secretory protein and components of the translocation machinery. The C<sub>ub</sub> moiety was placed 37 residues

downstream from the signal sequence of the  $\alpha$ -factor precursor, and the N terminus of Sec62p (see INTRODUCTION) was extended with  $N_{ub}$ . Using this version of the split-Ub assay, we could demonstrate that Sec62p is close to the nascent chain of the  $\alpha$ -factor during its translocation. By moving the  $C_{ub}$  moiety farther downstream from the signal sequence of the  $\alpha$ -factor precursor, we obtained "snapshots" of the relative proximity between the nascent polypeptide chain and Sec62p *in vivo*. The proximity thus detected was considerably reduced once the spacer sequence between the signal sequence of the  $\alpha$ -factor precursor and  $C_{ub}$  was increased from 37 to 65 residues (Figures 5B and 7B). The data strongly suggest that the access of Sec62p to the nascent chain of  $\alpha$ -factor is confined to a region of the nascent chain that is very close to the signal sequence. Similar results were obtained with the signal sequence of invertase as well. This property of Sec62p is the expected feature of a component of a signal sequence receptor.

Our interpretation, supported by several control experiments, is in agreement with the results produced by cross-linking and binding studies in cell-free systems (Musch *et al.*, 1992; Lyman and Schekman, 1997; Matlack *et al.*, 1997). Sec62p could be cross-linked to the  $\alpha$ -factor precursor *in vitro*, but only when ATP was omitted and the initiation of translocation of  $\alpha$ -factor was halted. Upon the addition of ATP, the translocation resumed and cross-linking was no longer possible (Musch *et al.*, 1992; Lyman and Schekman, 1997). The cross-linking between Sec62p and the nascent polypeptide chain was not observed when the translocating chain was halted in the ER channel (Musch *et al.*, 1992; Sanders *et al.*, 1992). It was therefore assumed that Sec62p is not a part of the channel and that it functions in the early steps of substrate recognition and initiation of translocation. The split-Ub assay, in its current form, depends on both halves of Ub being in the cell's cytosol. Therefore, the absence of the diagnostic cleavage (Figures 5B, 5C, and 7) when the  $C_{ub}$  moiety was placed farther downstream from the signal sequence (see RESULTS), while consistent with the absence of interactions between Sec62p and the nascent chain after the initiation of its translocation, does not address this issue directly. The  $C_{ub}$  moiety that emerges from the ribosome after it has docked at the ER channel is not accessible to the cytosolic  $N_{ub}$  moiety even if an  $N_{ub}$ -linked protein is spatially close to the translocation pore. This also explains the inability of  $N_{ub}$ -Sec61p to induce the cleavage of  $C_{ub}$ -containing translocation substrates bearing long spacer sequences between the signal sequence and  $C_{ub}$ , although the *in vitro* cross-linking studies have shown Sec61p to be in constant contact with the translocating polypeptide (Mothes *et al.*, 1994) (Figure 7).

The proximity between Sec62p (or Sec61p) and a translocating polypeptide is short lived. The rapid

transfer of the  $C_{ub}$  moiety into the lumen of the ER was shown to either prevent or strongly inhibit its interaction with the  $N_{ub}$  moiety of  $N_{ub}$ -Sec62p and  $N_{ub}$ -Sec61p (Figures 3A and 6C). In these experiments, the  $N_{ub}$  moieties bore either the glycine ( $N_{ug}$ ) or the alanine ( $N_{ua}$ ) residue at position 13 of  $N_{ub}$ . These modifications decrease the affinity between the two halves of Ub (see INTRODUCTION), thereby making the reconstitution of a quasi-native Ub moiety more dependent on the stability (half-life) of interactions between the proteins linked to  $N_{ub}$  and  $C_{ub}$ . These assays clearly distinguished the Sec62p-Sec61p signal sequence interactions from those that underlie the better understood, longer-lived protein complexes. For example, when linked to homodimerizing leucine zippers, the  $N_{ua}$  moiety, and even the  $N_{ug}$  moiety, is sufficient for reconstitution of the Ub (Johnsson and Varshavsky, 1994a).

#### *Is Sec62p Part of a Signal Sequence Receptor?*

The split-Ub assay measures the concentrations of the protein-coupled  $N_{ub}$  and  $C_{ub}$  moieties in the immediate vicinity of each other. Therefore, a positive result of a split-Ub assay signifies a spatial proximity between the two proteins but cannot, by itself, prove their physical interaction or address the functional significance of this proximity.  $N_{ub}$ -Sec62( $\Delta C60$ )Dha is functionally inactive and was shown to be not close to the translocating  $Mf\alpha_{37}$ - $C_{ub}$ -Dha (Figures 6 and 7). Since the sequence between  $N_{ub}$  and the first membrane-spanning region of Sec62 was retained in the C-terminally truncated Sec62( $\Delta C60$ )Dha, the distance between  $N_{ub}$  and the ER membrane was, most probably, not altered relative to wild-type Sec62p. The lack of significant cleavage of  $Mf\alpha_{37}$ - $C_{ub}$ -Dha in the presence of  $N_{ub}$ -Sec62( $\Delta C60$ )Dha must therefore result from the increased distance between Sec62( $\Delta C60$ )p and the translocating polypeptide chain. The C-terminal domain of intact Sec62p may contact other components of the translocation complex; alternatively, it may contribute to a binding site for the signal sequence or the nascent chain. These and related uncertainties notwithstanding, our results (Figures 6 and 7) provide the first *in vivo* evidence to support the view that Sec62p is part of a signal sequence-binding complex.

#### *Sec62p Discriminates between Different Signal Sequences*

The split-Ub assay has made it possible to show that Sec62p discriminates, in living cells, between two distinct signal sequences in otherwise identical fusion proteins. In the presence of  $N_{ub}$ -Sec62p, more of the free Dha reporter protein was produced *in vivo* from  $Mf\alpha_{37}$ - $C_{ub}$ -Dha than from  $Suc2_{23}$ - $C_{ub}$ -Dha (Figures 5 and 6). This selectivity is a property of Sec62p and not

a feature of the assay used, since approximately equal amounts of the cleaved reporter were produced with different signal sequences if the  $N_{ub}$  moiety was present as the  $N_{ub}$ -Sec61p fusion (Figures 6 and 7). This result confirmed, *in vivo*, that different targeting pathways converge at the Sec61-containing complex to initiate translocation. The existence of at least two different targeting pathways to the translocation pore was suggested by Walter and co-workers on the basis of the properties of yeast mutants that lacked either SRP or its receptor (Hann and Walter, 1991; Ogg *et al.*, 1992). One possibility is that the targeting via SRP operates cotranslationally, whereas the targeting via the Sec62/63 complex is predominantly posttranslational. The bifurcation between the cotranslational and posttranslational targeting is expected to be stochastic for many translocated proteins. Nonetheless, certain signal sequences do prefer SRP, while some of the other signal sequences are targeted by the Sec62/Sec63 complex (Ng *et al.*, 1996). Genetic studies have shown that the translocation of invertase continues in the presence of mutations in either the SRP or the Sec62/63 complex (Deshaias and Schekman, 1989; Ogg *et al.*, 1992). However, the kinetics and efficiency of invertase translocation are altered in the absence of SRP (Hann and Walter, 1991; Johnsson and Varshavsky, 1994b).

To explain the different efficiencies of cleavage of the two signal sequence-bearing  $C_{ub}$  fusions, we propose the following model. The signal sequence of invertase is recognized primarily by SRP and then transferred to the trimeric Sec61p complex, which completes the protein's translocation across the ER membrane. Under conditions that result in a shortage of SRP or a competition among different signal sequences, the Sec62p/Sec63p-containing complex, being a part of the alternative targeting pathway, would recognize an increasing fraction of invertase. This would explain why specific interactions between the invertase signal sequence-bearing proteins and Sec62p can only be observed for the more highly expressed  $C_{ub}$ -Dha fusions (Figure 5C). By contrast, the targeting of proteins bearing the signal sequence of the  $\alpha$ -factor precursor is mediated, *in vivo*, predominantly by the Sec62p/Sec63p-containing complex. This would account for the observed close proximity of these test proteins to Sec62p (Figures 5–7). If so, the split-Ub technique makes it possible to estimate the flux of two different secretory proteins through the targeting pathways to the ER membrane without the necessity of deleting or otherwise inactivating specific components of the targeting complex.

#### *Further Applications of the Split-Ub Technique*

The split-Ub sensor should also be applicable to other settings that involve short-lived protein interactions

that occur in the cytosol and are freely accessible to the Ub-specific proteases. The advantage of using this method for the analysis of protein translocation stems, in part, from the fast and irreversible removal of the translocated chain from the location (cytosol) where the  $N_{ub}/C_{ub}$  interaction is monitored. Similar situations are expected for the translocation of proteins into other organelles such as the mitochondrion, the nucleus, and the peroxisome.

The demonstration, in the present work, that Ura3p can serve as a reporter in a split-Ub assay (Figure 7) opens the way to genetic screens based on this assay. For example, introducing a DNA library consisting of random  $N_{ub}$ -gene fusions into a strain expressing a signal sequence-bearing  $C_{ub}$ -Ura3p protein should allow the identification of genes involved in targeting or translocation by enabling the cells to form colonies on media lacking uracil. This selection for protein ligands that interact transiently in the vicinity of a membrane complements the recent split-Ub-based screen for ligands that form relatively stable complexes (Stagljar *et al.*, 1998).

#### ACKNOWLEDGMENTS

We thank Ray Deshaias, Jürgen Dohmen, Nicole Lewke, Randy Schekman, and Sandra Wittke for the gifts of yeast strains and plasmids. M.D. and N.J. thank Silke Müller for excellent technical assistance. This work was supported by a grant to N.J. from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (0311107), and a grant to A.V. from the National Institutes of Health (GM-31530).

#### REFERENCES

- Allison, D.S., and Young, E.T. (1988). Single-amino-acid substitutions within the signal sequence of yeast prepro- $\alpha$ -factor affect membrane translocation. *Mol. Cell. Biol.* 8, 1915–1922.
- Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J., and Karin, M. (1997). Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Mol. Cell. Biol.* 17, 3094–3102.
- Banfield, D.K., Lewis, M.J., Rabouille, C., Warren, G., and Pelham, H.R. (1994). Localization of Sed5, a putative vesicle targeting molecule, to the *cis*-Golgi network involves both its transmembrane and cytoplasmic domains. *J. Cell Biol.* 127, 357–371.
- Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* 278, 2123–2126.
- Biederer, T., Volkwein, C., and Sommer, T. (1996). Degradation of subunits of the Sec61p complex, an integral component of the ER membrane, by the ubiquitin-proteasome pathway. *EMBO J.* 15, 2069–2076.
- Bird, P., Gething, M.J., and Sambrook, J. (1987). Translocation in yeast and mammalian cells: not all signal sequences are functionally equivalent. *J. Cell Biol.* 105, 2905–2914.
- Brodsky, J.L., and Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteoliposome. *J. Cell Biol.* 123, 1355–1363.

- Connolly, T., Collins, P., and Gilmore, R. (1989). Access of proteinase K to partially translocated nascent polypeptides in intact and detergent-solubilized membranes. *J. Cell Biol.* **108**, 299–307.
- Crowley, K.S., Liao, S., Worrell, V.E., Reinhart, G.D., and Johnson, A.E. (1994). Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. *Cell* **78**, 461–471.
- Crowley, K.S., Reinhart, G.D., and Johnson, A.E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. *Cell* **73**, 1101–1115.
- Deshaines, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* **349**, 806–808.
- Deshaines, R.J., and Schekman, R. (1989). SEC62 encodes a putative membrane protein required for protein translocation into the yeast endoplasmic reticulum. *J. Cell Biol.* **109**, 2653–2664.
- Deshaines, R.J., and Schekman, R. (1990). Structural and functional dissection of Sec62p, a membrane-bound component of the yeast endoplasmic reticulum protein import machinery. *Mol. Cell. Biol.* **10**, 6024–6035.
- Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. *J. Biol. Chem.* **270**, 18099–18109.
- Feldheim, D., and Schekman, R. (1994). Sec72p contributes to the selective recognition of signal peptides by the secretory polypeptide translocation complex. *J. Cell Biol.* **126**, 935–943.
- Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246.
- Gilmore, R., and Blobel, G. (1985). Translocation of secretory proteins across the microsomal membrane occurs through an environment accessible to aqueous perturbants. *Cell* **42**, 497–505.
- Görlich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* **71**, 489–503.
- Hanein, D., Matlack, K.E., Jungnickel, B., Plath, K., Kalies, K.U., Miller, K.R., Rapoport, T.A., and Akey, C.W. (1996). Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **87**, 721–732.
- Hann, B.C., and Walter, P. (1991). The signal recognition particle in *S. cerevisiae*. *Cell* **67**, 131–144.
- Hiller, M.M., Finger, A., Schweiger, M., and Wolf, D.H. (1996). ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. *Science* **273**, 1725–1728.
- Johnsson, N., and Varshavsky, A. (1994a). Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* **91**, 10340–10344.
- Johnsson, N., and Varshavsky, A. (1994b). Ubiquitin-assisted dissection of protein transport across membranes. *EMBO J.* **13**, 2686–2698.
- Johnsson, N., and Varshavsky, A. (1997). Split ubiquitin: a sensor of protein interactions in vivo. In: *The Yeast Two Hybrid System*, ed. P.L. Bartel and S. Fields, Oxford, United Kingdom: Oxford University Press, 316–332.
- Jungnickel, B., and Rapoport, T.A. (1995). A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. *Cell* **82**, 261–270.
- Krieg, U.C., Walter, P., and Johnson, A.E. (1986). Photocrosslinking of the signal sequence of nascent proprolactin to the 54-kilodalton polypeptide of the signal recognition particle. *Proc. Natl. Acad. Sci. USA* **83**, 8604–8608.
- Kurzchalia, T.V., Wiedmann, M., Girshovich, A.S., Bochkareva, E.S., Bielka, H., and Rapoport, T.A. (1986). The signal sequence of nascent proprolactin interacts with the 54K polypeptide of the signal recognition particle. *Nature* **320**, 634–636.
- Lyman, S.K., and Schekman, R. (1997). Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell* **88**, 85–96.
- Maarse, A.C., Blom, J., Grivell, L.A., and Meijer, M. (1992). MPI1, an essential gene encoding a mitochondrial membrane protein, is possibly involved in protein import into yeast mitochondria. *EMBO J.* **11**, 3619–3628.
- Matlack, K.E., Plath, K., Misselwitz, B., and Rapoport, T.A. (1997). Protein transport by purified yeast Sec complex and Kar2p without membranes. *Science* **277**, 938–941.
- Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. (1997). Fluorescent indicators for  $\text{Ca}^{2+}$  based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–887.
- Mothes, W., Prehn, S., and Rapoport, T.A. (1994). Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. *EMBO J.* **13**, 3973–3982.
- Müsch, A., Wiedmann, M., and Rapoport, T.A. (1992). Yeast Sec proteins interact with polypeptides traversing the endoplasmic reticulum membrane. *Cell* **69**, 343–352.
- Nakano, A., Brada, D., and Schekman, R. (1988). A membrane glycoprotein, Sec12p, required for protein transport from the endoplasmic reticulum to the Golgi apparatus in yeast. *J. Cell Biol.* **107**, 851–863.
- Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* **134**, 269–278.
- Ogg, S.C., Poritz, M.A., and Walter, P. (1992). Signal recognition particle receptor is important for cell growth and protein secretion in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **3**, 895–911.
- Orlean, P., Kuranda, M.J., and Albright, C.F. (1991). Analysis of glycoproteins from *Saccharomyces cerevisiae*. *Methods Enzymol.* **194**, 682–697.
- Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport, T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561–570.
- Plempert, R.K., Böhmler, S., Bordallo, J., Sommer, T., and Wolf, D.H. (1997). Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. *Nature* **388**, 891–895.
- Rapoport, T.A., Jungnickel, B., and Kutay, U. (1996). Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303.
- Rossi, F., Charlton, C.A., and Blau, H.M. (1997). Monitoring protein-protein interactions in intact eukaryotic cells by  $\beta$ -galactosidase complementation. *Proc. Natl. Acad. Sci. USA* **94**, 8405–8410.
- Rothblatt, J.A., Deshaines, R.J., Sanders, S.L., Daum, G., and Schekman, R. (1989). Multiple genes are required for proper insertion of secretory proteins into the endoplasmic reticulum in yeast. *J. Cell Biol.* **109**, 2641–2652.
- Sanders, S.L., Whitfield, K.M., Vogel, J.P., Rose, M.D., and Schekman, R.W. (1992). Sec61p and BiP directly facilitate polypeptide translocation into the ER. *Cell* **69**, 353–365.
- Sapperstein, S., Berkower, C., and Michaelis, S. (1994). Nucleotide sequence of the yeast STE14 gene, which encodes farnesylcysteine

- carboxyl methyltransferase, and demonstration of its essential role in a-factor export. *Mol. Cell. Biol.* 14, 1438–1449.
- Shim, J., Newman, A.P., and Ferro-Novick, S. (1991). The BOS1 gene encodes an essential 27-kDa putative membrane protein that is required for vesicular transport from the ER to the Golgi complex in yeast. *J. Cell Biol.* 113, 55–64.
- Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122, 19–27.
- Simon, S.M., and Blobel, G. (1991). A protein-conducting channel in the endoplasmic reticulum. *Cell* 65, 371–380.
- Stagljar, I., Korostensky, C., Johnsson, N., and te Heesen, S. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc. Natl. Acad. Sci. USA* 95, 5187–5192.
- Walter, P., and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* 10, 87–119.
- Walter, P., Ibrahim, I., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum I. Signal recognition protein (SRP) binds to in vitro assembled polysomes synthesizing secretory protein. *J. Cell Biol.* 91, 545–550.
- Wiertz, E.J., Tortorella, D., Bogyo, M., Yu, J., Mothes, W., Jones, T.R., Rapoport, T.A., and Ploegh, H.L. (1996). Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature* 384, 432–438.
- Wilkinson, B.M., Critchley, A.J., and Stirling, C.J. (1996). Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. *J. Biol. Chem.* 271, 25590–25597.

## ORIGINAL PAPER

Cristina Pungartnik · Marcelo F. Kern  
Martin Brendel · João A. P. Henriques

## Mutant allele *pso7-1*, that sensitizes *Saccharomyces cerevisiae* to photoactivated psoralen, is allelic with *COX11*, encoding a protein indispensable for a functional cytochrome c oxidase

Received: 18 May / 8 June 1999

**Abstract** The yeast gene *PSO7* was cloned from a genomic library by complementation of the *pso7-1* mutant's sensitivity phenotype to 4-nitroquinoline-1-oxide (4NQO). Sequence analysis revealed that *PSO7* is allelic to the 1.1-kb ORF of the yeast gene *COX11* which is located on chromosome XVI and encodes a protein of 28-kDa localized in the inner mitochondrial membrane. Allelism of *PSO7/COX11* was verified by non-complementation of 4NQO-sensitivity in diploids homo- and hetero-allelic for the *pso7-1* and *cox11::TRP1* mutant alleles. Sensitivity to 4NQO was the same in exponentially growing cells of the *pso7-1* mutant and the *cox11::TRP1* disruptant. Allelism of *COX11* and *PSO7* indicates that the *pso7* mutant's sensitivity to photoactivated 3-carbethoxypsoralen and to 4NQO is not caused by defective DNA repair, but rather is due to an altered metabolism of the pro-mutagen 4NQO in the absence of cytochrome oxidase (Cox) in *pso7-1/cox11::TRP1* mutants/disruptants. Lack of Cox might also lead to a higher reactivity of the active oxygen species produced by photoactivated 3-carbethoxypsoralen. The metabolic state of the cells is important for their sensitivity phenotype since the largest enhancement of sensitivity to 4NQO between wild-type (WT) and the *pso7* mutant occurs in exponentially growing cells, while cells in stationary phase or growing cells in phosphate buffer have the same 4NQO resistance, irrespective of their WT/mutant status. Strains containing the *pso7-1* or *cox11::TRP1* mutant allele were also sensitive to the

oxidative stress-generating agents H<sub>2</sub>O<sub>2</sub> and paraquat. Mutant *pso7-1*, as well as disruptant *cox11::TRP1*, harboured mitochondria that in comparison to WT contained less than 5% and no detectable Cox activity, respectively.

**Key words** Psoralen sensitivity · Cytochrome oxidase · *Saccharomyces cerevisiae* · Oxidative stress

### Introduction

The *pso* mutants of *Saccharomyces cerevisiae* are sensitive to photoactivated psoralens. *PSO* genes have recently been divided into two groups; those which seem directly involved in DNA repair and those that most-likely are not. Whereas *PSO1* (allelic to *REV3*), *PSO2* (allelic to *SNM1*), *PSO3*, *PSO4* (allelic to *PRP19*), and *PSO5* (allelic to *RAD16*) encode proteins with apparent functions in error-free and error-prone repair (see review by Henriques et al. 1997) the genes *PSO6* and *PSO7*, by virtue of the weak sensitivity phenotype of their mutant alleles (Querol et al. 1994), are thought not to represent genes encoding repair proteins. Sensitivity in these latter mutants is mainly for treatments involving agents that, amongst other features, induce oxidative stress (Querol et al. 1994). *PSO6* has recently been cloned and found to be allelic with the gene *ERG3* encoding sterol C-5 desaturase, an enzyme necessary for the synthesis of ergosterol (Schmidt et al. 1999).

For *PSO7* the only existing mutant allele, *pso7-1*, was shown to primarily confer a moderate sensitivity to photoactivated 3-carbethoxypsoralen (3CPs + UVA). Stationary phase cells of the *pso7-1* mutant have nearly wild-type (WT) resistance for a range of mutagens, namely UV<sub>254 nm</sub>, 8-MOP + UVA, nitrogen mustard (HN1), and 4NQO (Querol et al. 1994). However, *pso7-1* mutants do exhibit increased sensitivity to 4NQO when the pro-mutagen is added to exponentially growing cells. This feature was used in our strategy to molecularly clone *PSO7* via the isolation of 4NQO-resistant

C. Pungartnik · M.F. Kern · J.A.P. Henriques (✉)  
Departamento de Biofísica/Centro de Biotecnologia, UFRGS,  
Av. Bento Gonçalves 9500,  
CEP 91507-970 Porto Alegre, RS, Brasil  
e-mail: pegas@dna.cbiof.ufrgs.br

M. Brendel  
Institut für Mikrobiologie der J.W. Goethe-Universität,  
Theodor-Stern-Kai 7, Hs. 75  
60590 Frankfurt am Main, Germany

Communicated by K. Wolf

- Singer-Kruger B, Ferro-Novick S (1997) Use of a synthetic lethal screen to identify yeast mutants impaired in endocytosis, vacuolar protein sorting and the organization of the cytoskeleton. *Eur J Cell Biol* 74:365–375
- Solomon V, Baracos V, Sarraf P, Goldberg A (1998 a) Rates of ubiquitin conjugation increase when muscles atrophy, largely through activation of the N-end rule pathway. *Proc Natl Acad Sci USA* 95:12602–12607
- Solomon V, Lecker SH, Goldberg AL (1998 b) The N-end rule pathway catalyzes a major fraction of the protein degradation in skeletal muscle. *J Biol Chem* 273:25216–25222
- Taban CH, Hondermarck H, Bradshaw RA, Boilly B (1996) Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. *Experientia* 52:865–870
- Tanaka J, Fink GR (1985) The histidine permease gene (*HIP1*) of *Saccharomyces cerevisiae*. *Gene* 38:205–214
- Tobias JW, Shrader TE, Rocap G, Varshavsky A (1991) The N-end rule in bacteria. *Science* 254:1374–1377
- Varshavsky A (1996) The N-end rule: functions, mysteries, uses. *Proc Natl Acad Sci USA* 93:12142–12149
- Varshavsky A (1997) The ubiquitin system. *Trends Biochem Sci* 22:383–387
- Varshavsky A, Byrd C, Davydov IV, Dohmen RJ, Du F, Ghislain M, Gonzalez M, Grigoryev S, Johnson ES, Johnsson N, Johnston JA, Kwon YT, Lévy F, Lomovskaya O, Madura K, Ota I, Rümenapf T, Shrader TE, Suzuki T, Turner G, Waller PRH, Webster A (1998) The N-end rule pathway. In: Peters J-M, Harris JR, Finley D (eds) *Ubiquitin and the biology of the cell*. Plenum Press, New York, pp 223–278
- Worley CK, Ling R, Callis J (1998) Engineering *in vivo* instability of firefly luciferase and *Escherichia coli* beta-glucuronidase in higher plants using recognition elements from the ubiquitin pathway. *Plant Mol Biol* 37:337–347

## ORIGINAL PAPER

Youming Xie · Alexander Varshavsky

## The N-end rule pathway is required for import of histidine in yeast lacking the kinesin-like protein Cin8p

Received: 7 April / 17 May 1999

**Abstract** The N-end rule pathway is a ubiquitin-dependent proteolytic system whose targets include proteins bearing destabilizing N-terminal residues. We carried out a synthetic lethal screen for *Saccharomyces cerevisiae* mutants that require the N-end rule pathway for cell viability. A mutant thus identified, termed *sln2*, could not grow in the absence of Ubr1p, the recognition component of the N-end rule pathway, which was not essential for viability of the parental strain under the same conditions. Further analysis showed that inviability of *sln2 ubr1Δ* cells could be rescued either by the *HIS3* gene (which was absent from the parental strain) or by a high concentration of histidine in the medium. This defect in histidine uptake, exhibited by the *sln2* mutant in the absence but not in the presence of Ubr1p, was traced to the gene *HIP1*, which encodes the histidine transporter. *HIP1* was underexpressed in *sln2 ubr1Δ* cells, in comparison to either *sln2 UBR1* or *SLN2 ubr1Δ* cells. Yet another property of the *sln2* mutant was its inviability at 37 °C, which could not be rescued by either *UBR1* or *HIS3*. This feature of *sln2* allowed the cloning of *SLN2*, which was found to be a gene called *CIN8*, encoding a kinesin-like protein. Thus, either the N-end rule pathway or Cin8p must be present for the viability-sustaining rate of histidine import in *S. cerevisiae* auxotrophic for histidine. We consider possible mechanisms of this previously unsuspected link between kinesins, ubiquitin-dependent proteolysis, and the import of histidine.

**Key words** Ubiquitin · Proteolysis · Kinesin · Histidine import · *CIN8* · *HIP1* · *UBR1*

### Introduction

A number of regulatory circuits involve metabolically unstable proteins. A short *in vivo* half-life of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration, or subunit composition, through changes in the rate of regulator's synthesis or degradation. Short *in vivo* half-lives are also characteristic of damaged or otherwise abnormal proteins (Hochstrasser 1996; Varshavsky 1997; Peters et al. 1998; Scheffner et al. 1998). Features of proteins that confer metabolic instability are called degradation signals, or degrons. The essential component of one degradation signal, the first to be identified, is a destabilizing N-terminal residue of a protein (Bachmair et al. 1986; Varshavsky 1996). The set of amino-acid residues which are destabilizing in a given cell yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct N-end rule pathways are present in all organisms examined, from mammals and plants to fungi and bacteria (Bachmair and Varshavsky 1989; Gonda et al. 1989; Tobias et al. 1991; Varshavsky 1996; Worley et al. 1998).

In eukaryotes, the degradation signals (N-degrons) recognized by the N-end rule pathway comprise two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (Bachmair and Varshavsky 1989; Hill et al. 1993; Dohmen et al. 1994). The Lys residue is the site of formation of a multiubiquitin chain (Chau et al. 1989). The N-end rule pathway is thus one pathway of the ubiquitin system. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (Johnson et al. 1995; Hochstrasser 1996; Hershko 1997; Pickart 1997; Varshavsky 1997; Scheffner et al. 1998). In most of these processes, Ub acts through routes that involve the degradation of ubiquitin-protein conjugates by the 26S proteasome, an ATP-dependent

Y. Xie · A. Varshavsky (✉)  
Division of Biology, 147-75, Caltech, 1200 East California Blvd.,  
Pasadena, CA 91125, USA  
e-mail: avarsh@cco.caltech.edu  
Tel.: +1-626-395 3785; Fax: +1-626-440 9821

Communicated by C.P. Hollenberg

multisubunit protease (Rechsteiner et al. 1993; Coux et al. 1996; Baumeister et al. 1998; Glickman et al. 1998).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are *tertiary* destabilizing N-terminal residues in that they function through their conversion into the *secondary* destabilizing N-terminal residues Asp and Glu. This conversion is mediated by the *NTA1*-encoded N-terminal amidase (Nt-amidase) (Baker and Varshavsky 1995). The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing residues (Balzi et al. 1990). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3), the recognition component of the N-end rule pathway (Bartel et al. 1990). In *S. cerevisiae*, N-recognin is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues – Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (Varshavsky 1996). The *Ubr1*-encoded mammalian N-recognin (E3 $\alpha$ ) has also been characterized (Hershko 1991; Kwon et al. 1998). Ubr1p has at least two substrate-binding sites. The type-1 site is specific for the basic N-terminal residues Arg, Lys and His. The type-2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (Varshavsky 1996). Both yeast and mammalian Ubr1p have also been shown to recognize a set of internal degrons that remain to be characterized (Gonen et al. 1991; Madura and Varshavsky 1994; Byrd et al. 1998).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2* which encodes the peptide transporter (Alagramam et al. 1995; Byrd et al. 1998); a role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (Ota and Varshavsky 1993); the degradation of Gpa1p, a G $\alpha$  protein of *S. cerevisiae* (Madura 1994, #60); and the conditional degradation of alphavirus RNA polymerase in virus-infected metazoan cells (deGroot et al. 1991; Varshavsky et al. 1998). Physiological N-end rule substrates were also identified among the proteins secreted into the mammalian cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes* (Sijts et al. 1997). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (Hondermarck et al. 1992), and to delay limb regeneration in amphibians (Taban et al. 1996). In mammals, the N-end rule pathway is also likely to play a role in catabolic states that result in muscle atrophy (Solomon et al. 1998 a,b).

While the absence of the N-end rule pathway (specifically, the absence of Ubr1p) from *S. cerevisiae* is lethal under conditions where the import of peptides is essential for cell survival (Alagramam et al. 1995; Byrd et al. 1998), *ubr1Δ* cells are viable and only slightly

growth-retarded on standard media (Bartel et al. 1990). Synthetic lethal screens can often identify pathways whose essential functions overlap (or interact) with the function of a system of interest (Singer-Kruger and Ferro-Novick 1997; Ho et al. 1998; Merrill and Holm 1998). An earlier screen for *S. cerevisiae* mutants that require Ubr1p for viability has identified Sln1p, which contains a sensor kinase and response regulator domains analogous to those of the prokaryotic two-component regulators (Ota and Varshavsky 1993). Sln1p functions in the HOG signaling pathway, which regulates the response of cells to changes in the media osmolarity (Posas et al. 1996). *sln1* mutants are growth-impaired in presence of Ubr1p and inviable in its absence (Ota and Varshavsky 1993). The identity of a relevant Ubr1p substrate in the HOG pathway is unknown.

In the present work, we carried out a modified synthetic lethal screen for *S. cerevisiae* mutants that require Ubr1p for viability. Our findings uncovered a previously unsuspected connection between kinesins, the N-end rule pathway, and the import of histidine.

## Materials and methods

**Yeast strains, media, and genetic procedures.** *S. cerevisiae* strains used in this study are listed in Table 1. Yeast were grown in rich (YPD) medium, or in synthetic media containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients, including adenine, arginine, histidine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and uracil, and either 2% dextrose (glucose) (SD medium) or 2% galactose (SG medium) (Sherman et al. 1986). In some experiments, the concentration of histidine in SD medium was increased 10-fold, to 0.2 mg/ml. Unless stated otherwise, cells were incubated at 30 °C. Yeast mating, sporulation, and tetrad dissection were performed as described (Sherman et al. 1986). The mating-type switch was carried out through transformation with the plasmid pGAL-HO (Herskowitz and Jensen 1991). *S. cerevisiae* were transformed as described by Chen et al. (1992). The *Escherichia coli* strain used was DH5 $\alpha$  (Ausubel et al. 1996).

**Synthetic lethal screen.** The haploid *P<sub>GAL1</sub>-UBR1* strain JD54, in which the *UBR1* gene is expressed from the *P<sub>GAL1</sub>* promoter (Ghislain et al. 1996), was mutagenized with ethyl methane-sulfonate (EMS) (Sigma, St. Louis, Mo.) to approximately 7% survival upon plating to SG medium at 30 °C. The colonies on SG plates were replica-plated onto SD plates. The colonies that grew on SG but not on SD media were re-tested for the absence of growth on SD plates. The candidate isolates were transformed with plasmids that expressed *UBR1* either from its natural promoter [pUBR1 (Bartel et al. 1990)] or the much stronger *P<sub>ADH1</sub>* promoter, and tested for survival on SD plates. One mutant, termed *sln2* (AVY200), was rescued by *UBR1* on SD plates, and was selected for further study.

**Isolation of HIS3 as a suppressor of *sln2*.** The *sln2* mutant was transformed with a yeast genomic DNA library (American Type Culture Collection, #77164) in the *TRP1*, *CEN6*-based vector pRS200 (Sikorski and Hieter 1989). Cells were plated on SD(-Trp) plates and incubated at 30 °C for 2–3 days. Out of approximately 4 × 10<sup>4</sup> transformants, 25 could grow on SD(-Trp). Library plasmids were rescued from these transformants and re-transformed into *sln2* cells. Eight of these plasmids conferred viability on *sln2* cells after re-transformation and selection on SD(-Trp). Restriction analyses showed that these plasmids were either identical or had

**Table 1** *S. cerevisiae* strains used in this study

Strain	Genotype	References
JD52	<i>MATa ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112</i>	Johnson et al. (1995)
JD54	<i>MATa ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1</i>	Ghislain et al. (1996)
AVY199	<i>MATa ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1</i>	Derivative of JD54
AVY200 <sup>a</sup>	<i>MATa sln2 ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1</i>	Derivative of JD54
AVY201	<i>MATa ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1 hip1Δ-3::URA3</i>	Derivative of JD54
AVY202	<i>MATa ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1 cin8Δ-6::URA3</i>	Derivative of JD54
AVY203	<i>MATa ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 cin8Δ-10::URA3</i>	Derivative of JD52
AVY204	<i>MATa sln2 ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1 cup9Δ-1::LEU2</i>	Derivative of AVY200

<sup>a</sup> Mutant (*sln2*) that requires *UBR1* for viability

large fragments in common. Truncation and sequencing analyses localized the complementing activity to a 1.7-kb *Bam*H I fragment containing the *HIS3* gene. This finding was independently confirmed by showing that the *HIS3* open reading frame (ORF), expressed from the *P<sub>MET25</sub>* promoter (Mumberg et al. 1994), was also able to complement the *sln2* mutant on SD plates.

**Histidine uptake assay.** This assay was carried out as described (Grauslund et al. 1995), with minor modifications. Briefly, the strains to be assayed were grown to A<sub>600</sub> of about 0.6 in SD(-His) medium at 30°C. Cells were harvested from 5-ml cultures and resuspended in 0.4–0.5 ml of SD(-His) at an identical cell density, adjusted using a hemocytometer. For each sample, 0.1 ml of suspension was added to 0.1 ml of SD(-His) medium containing 8 μM of L-[<sup>14</sup>C]histidine (Amersham, Arlington Heights, Ill.; specific radioactivity 0.3 Ci/mmol), and incubated at 30°C for 2 min. Samples were removed, filtered through a #30 glass-fiber filter (Schleicher and Schuell), washed with 10 ml of ice-cold water, and air-dried. <sup>14</sup>C radioactivity was determined using a liquid scintillation counter. Since high concentrations of histidine in the medium would have interfered with the assay, *HIS3* was introduced (as the pRS313 plasmid) into all *his3Δ* strains to be tested, making them prototrophic for histidine.

**Northern hybridization.** *S. cerevisiae* strains were grown at 30°C to A<sub>600</sub> of about 1. RNA was isolated as described (Schmitt et al. 1990) and fractionated by electrophoresis in formaldehyde-containing agarose gels (Ausubel et al. 1996). Briefly, approximately 40 μg of RNA per lane was fractionated in a 1.2% agarose gel containing 0.66 M formaldehyde, 1 × MOPS buffer and 0.1 μg/ml of ethidium bromide (Ausubel et al. 1996). Electrophoresis was performed at 5 V/cm in 1 × MOPS buffer. Fractionated RNA was transferred onto a GeneScreen Plus membrane (DuPont NEN, Boston, Mass.) according to the manufacturer's instructions. DNA probes were labeled with [ $\alpha$ -<sup>32</sup>P]dCTP using a DNA labeling kit (Pharmacia Biotech, Piscataway, N.J.). Hybridization was carried out using a NorthernMax kit (Ambion, Austin, Tex.). Hybridization patterns were detected by autoradiography and quantitated using PhosphorImager (Molecular Dynamics).

**hip1Δ mutant.** This mutant was produced from JD54 (Table 1) using a direct deletion technique described previously (Baudin et al. 1993). Briefly, two oligodeoxynucleotides of 66 nt were synthesized, attagtacacaatacagatatacaacatcaaaaatgtcgaaagtcatacatataaggaaacg and gatatggaaattccgcgtatagctcatctttcccttcatttttgcgtgccttcatcttc. One oligonucleotide comprised 40 nt of a sequence derived from the 5' untranslated region (UTR) of the *HIP1* gene, followed by 26 nt of a sequence derived from the *URA3* ORF, beginning from its start codon. The second oligonucleotide

comprised 40 nt of a sequence derived from the 3' UTR of *HIP1*, followed by 26 nt of a sequence derived from the 3'-proximal region of the *URA3* ORF, including its stop codon. The two oligonucleotides were used as PCR primers to amplify *URA3* from the pRS306 vector (Sikorski and Hieter 1989). The resulting targeting fragment contained *URA3* flanked on either side by 40-bp sequences derived from, respectively, the 5' and 3' UTRs of *HIP1*. JD54 cells (Table 1) were transformed with the targeting fragment, and selected for Ura<sup>+</sup> transformants at 30°C on SD(-Ura) plates containing 0.2 mg/ml of histidine. In a typical experiment, approximately 50% of Ura<sup>+</sup> transformants were the expected *hip1Δ* mutants, as verified by PCR, and also by checking for the phenotype of dependence on high concentrations of histidine in the medium.

**Isolation of *HIP1*.** A gap-repair method was used to rescue *HIP1* from both wild-type (JD54) and *sln2* (AVY200) strains (Rothstein 1991). Two fragments (the left and right arms) flanking the gap were amplified by PCR from either JD54 or *sln2* genomic DNA. The left arm was a 693-bp *Xba*I-*Hind*III fragment corresponding to the 5' UTR of *HIP1* from -976 bp to -284 bp. The right arm was a 1044-bp *Hind*III-*Xba*I fragment, corresponding to the 3' UTR of *HIP1*, 470 bp downstream from the stop codon of *HIP1*. The left and right arms were inserted into the *Xba*I and *Xba*I sites of pRS315, a *LEU2*, *CEN6*-based vector (Sikorski and Hieter 1989). The resulting construct was linearized with *Hind*III, and transformed into wild-type and *sln2* cells (see Fig. 3 A). Plasmids were isolated from the Leu<sup>+</sup> transformants. Gap-filled plasmids (pRS315HIP1) were identified by restriction analysis, and also by sequencing the relevant junctional regions. The primers for producing the left-arm fragment were ggggctgaaacacaagaatcataatg and gcacctactcaactcttagcag. The initial PCR product was digested with *Xba*I and *Hind*III, yielding the left-arm DNA fragment. The primers for producing the right arm were ctactacagatattacatgtggcg and gggttcggtagaaatcggttagattg. The initial PCR product was digested with *Hind*III and *Xba*I, yielding the right arm. All PCR products were verified by DNA sequencing.

**Construction of the *sln2 cup9Δ* mutant.** The previously described *cup9Δ::LEU2* disruption allele (Byrd et al. 1998) was used to replace the *CUP9* allele in *sln2* cells through homologous recombination (Rothstein 1991), yielding the strain AVY204 (Table 1). The absence of Cup9p was verified by testing for the ability of AVY204 cells to import dipeptides in the absence of *UBR1* (on SD plates) (Byrd et al. 1998), and also by PCR analysis.

**Isolation of the *SLN2* (*CIN8*) gene.** The *S. cerevisiae* genomic DNA library used for the isolation of *HIS3* (see above) was transformed into the *sln2* strain (AVY200) (Table 1). Cells were

plated on SD(-Trp) plates containing histidine at 0.2 mg/ml, and incubated at 30°C for 8 h. The temperature was then raised to 37°C, until the colonies formed (about 3 days). Out of approximately  $6 \times 10^4$  transformants at 30°C, 15 could grow on SD(-Trp, +0.2 mg/ml His) at 37°C. Library derived plasmids were rescued from these 15 transformants and re-transformed into *sln2* cells, with selection on SD(-Trp) plates at 37°C. Three plasmids conferred viability on *sln2* cells after re-transformation. Restriction mapping showed the inserts of these plasmids to be identical. Truncation and sequence analyses confirmed that the complementing activity was confined to the *CIN8* locus.

**Isolation of the *CIN8\** allele from *sln2* cells.** A gap-repair method (Rothstein 1991) described above for *HIP1* was used to isolate the mutant allele of *CIN8* (denoted as *CIN8\**) from *sln2* cells. A 4973-bp *EcoRI-BamHI* fragment containing wild-type *CIN8*, including its 5' (464 bp) and 3' (1353 bp) untranslated regions (UTRs), was subcloned into the *TRP1*-based vector pRS314 (Sikorski and Hieter 1989). The resulting pRS314CIN8 was gapped by digestion with *PacI* and *BseRI*, and transformed into the *sln2* strain (AVY200). Plasmids were isolated from the *Trp<sup>+</sup>* transformants. The gap-filled plasmids (pRS314CIN8\*) were verified by restriction analysis and sequencing of the relevant junctional regions.

**Construction of null *cin8* mutants.** An approximately 2.5-kb *SphI-AatII* DNA fragment of pRS314CIN8 was replaced by a PCR-produced, approximately 1.1-kb, URA3-containing *SphI-AatII* fragment of YIp5, yielding pRS314cin8::URA3. A *Sall-BamHI* fragment of pRS314cin8::URA3 containing the *cin8*::URA3 deletion/disruption allele was used to replace, through homologous recombination, the *CIN8* gene in JD52, a wild-type (*Ubr1<sup>+</sup>*) strain, and in JD54, a *P<sub>GAL1</sub>::UBR1* strain. The resulting *cin8Δ* mutants AVY203 and AVY202, respectively (Table 1), lacked codons 169–1015 of the *CIN8* ORF. The structure of the *cin8Δ* allele, and the temperature sensitivity of *cin8Δ* mutants were confirmed, respectively, using PCR and growth tests at 37°C. The primers for producing the *URA3*-containing DNA fragment were cattG-CATGcgccgtggaaatttcatgtttgac and agtgGACGTCTaccagggtgcct-gactgcgtttagc (uppercase letters denote the *SphI* and *AatII* sites, respectively).

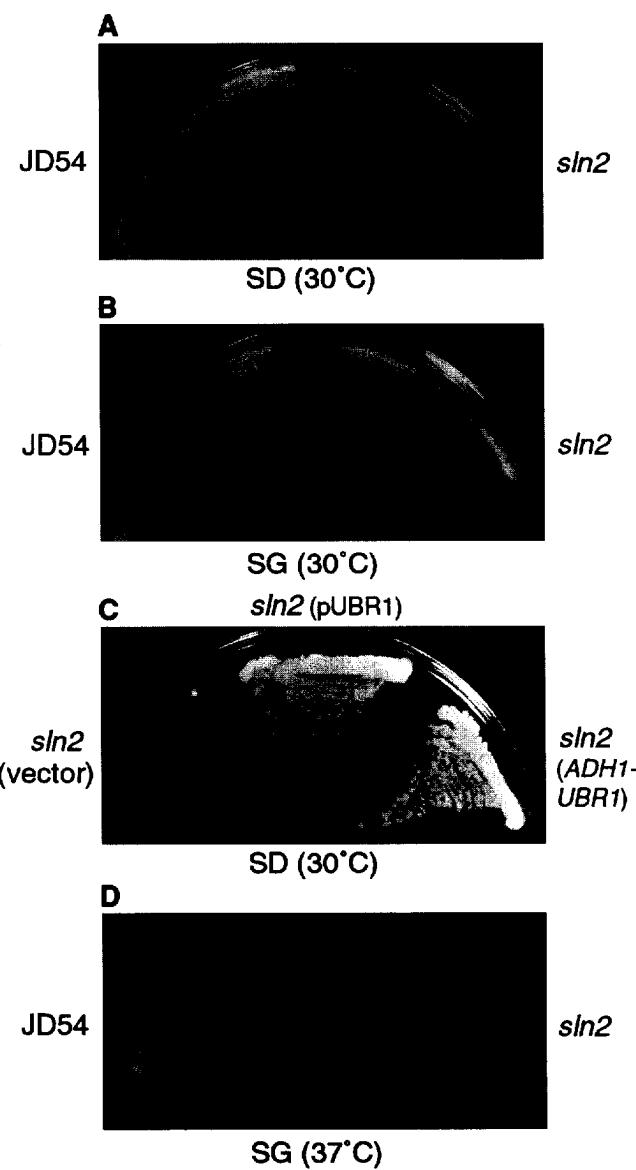
## Results

### Isolation of the *sln2* mutant

The earlier screen for *S. cerevisiae* mutants requiring the N-end rule pathway (specifically, *Ubr1p*) for viability (Ota and Varshavsky 1993) was far from exhaustive, and in addition utilized wild-type levels of *Ubr1p* expression. Since the activity of the N-end rule pathway is higher in cells overproducing *Ubr1p* (Bartel et al. 1990; Madura and Varshavsky 1994), the present screen was carried out using the strain JD54 (Ghislain et al. 1996), in which *UBR1* was expressed from the galactose-inducible, dextrose-repressible *P<sub>GAL1</sub>* promoter (Table 1). Mutagenized JD54 cells were plated onto SG (galactose-containing) plates, and the colonies formed were replicated onto SD (dextrose-containing) plates, where the N-end rule pathway was inactive, owing to the absence of *Ubr1p* expression. Cells that yielded colonies on SG but not on SD plates were putative *sln* (synthetic lethal of N-end rule) mutants. Four such isolates were identified among approximately  $3.5 \times 10^5$  colonies screened.

To verify that the conditional viability of these isolates was in fact dependent on the presence of *Ubr1p*,

rather than on the nature of carbon source (dextrose versus galactose), the putative mutants were transformed with a low-copy plasmid pUBR1 expressing *UBR1* from its natural promoter, and plated on SD plates. Only one isolate, termed *sln2* (strain AVY200; Table 1), formed colonies on SD plates in the presence of pUBR1 but not in the presence of the control vector (Fig. 1 A–C). The inviability of *sln2* cells on SD plates



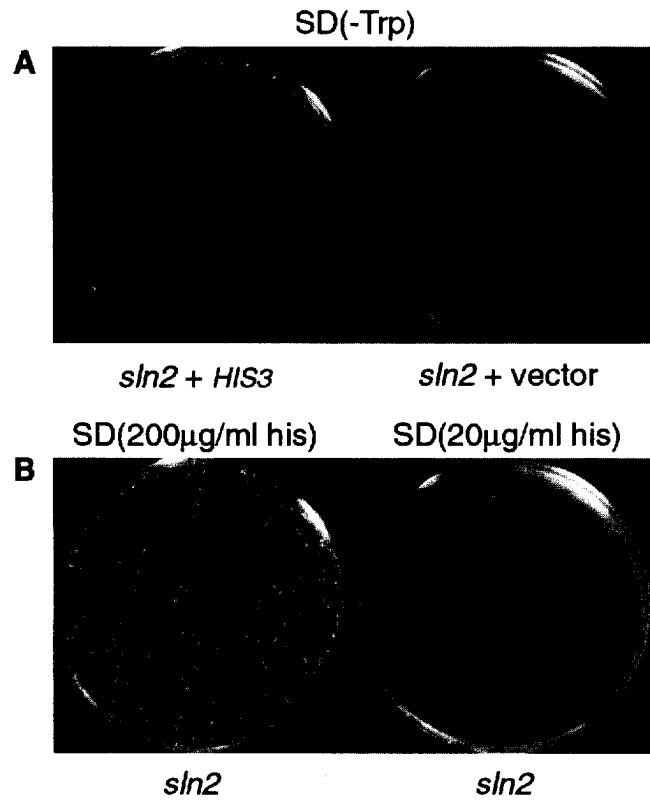
**Fig. 1 A–D** A *S. cerevisiae* mutant whose viability requires the N-end rule pathway. (A and B) the AVY200 (*sln2 P<sub>GAL1</sub>::UBR1*) strain grows on SG but not on SD plates (see Results). (C) evidence that the viability of *sln2* mutant requires *Ubr1p* rather than galactose. AVY200 (*sln2 P<sub>GAL1</sub>::UBR1*) cells were streaked on SD plates after having been transformed either with pUBR1, a low-copy plasmid expressing *UBR1* from its natural promoter or with pADH-UBR1, another *CEN*-based plasmid expressing *UBR1* from the constitutive and strong *P<sub>ADH1</sub>* promoter, or with vector alone. (D) *sln2* cells are *ts* for growth. The parental JD54 (*P<sub>GAL1</sub>::UBR1*) strain and the AVY200 (*sln2 P<sub>GAL1</sub>::UBR1*) strain (Table 1) were streaked on SG plates and incubated at 37°C

could be rescued by *UBR1* expressed at either normal (pUBR1) or much higher levels (pRB208, which expressed *UBR1* from the *P<sub>ADH1</sub>* promoter) (Fig. 1 C).

#### Histidine import defect of *sln2* cells in the absence of Ubr1p

Crossing AVY200 (*sln2*) to the strain AVY199 (congenic to JD54 but of the opposite mating type) (Table 1) indicated that the *sln2* mutation was recessive (data not shown). We used a library of *S. cerevisiae* genomic DNA fragments in a low-copy vector to transform AVY200 cells, screening for genes that could rescue the inviability of *sln2* on SD plates (in the absence of Ubr1p). Surprisingly, *HIS3*, a gene of the histidine biosynthesis pathway, was found to be a strong suppressor of the lethality of AVY200 (*sln2*) on SD plates (Fig. 2 A). Since AVY200 is a *his3Δ* strain (Table 1), *HIS3* is, by definition, an extragenic suppressor of the conditionally lethal *sln2* phenotype, i.e., *HIS3* is distinct from *SLN2*.

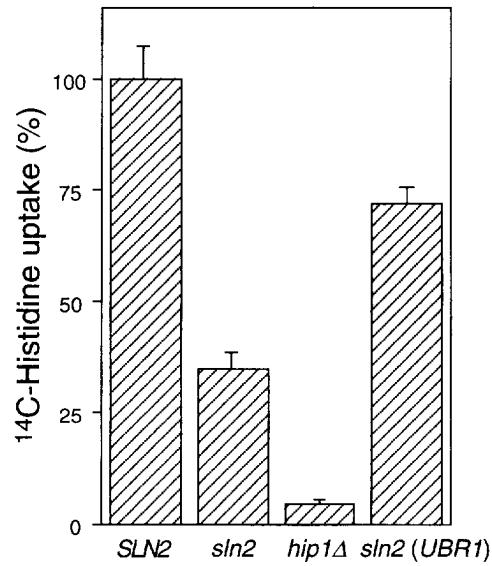
The original JD54 (*his3Δ*) strain, being a histidine auxotroph, required the presence of histidine in the medium. In the presence of *HIS3*, the JD54-derived AVY200 (*sln2*) mutant would be expected to become prototrophic for histidine. Could the effect of the *sln2* mutation be an impairment of the histidine uptake in the



**Fig. 2** Inviability of the AVY200 (*sln2 P<sub>GAL</sub>-UBR1*) strain on SD plates can be rescued either by the *HIS3* gene (panel A) or by high (>75  $\mu$ g/ml) concentration of histidine in the medium (panel B). See Results for details

absence of the N-end rule pathway (specifically, in the absence of Ubr1p)? This interpretation could account for the suppression of inviability of AVY200 cells on SD plates by the *HIS3* gene, inasmuch as the resulting strain would be a histidine prototroph. We first asked whether increasing the concentration of histidine in the growth medium would rescue AVY200 (*sln2*) cells on SD plates. [It has previously been shown that cells defective in histidine uptake could grow on a medium containing sufficiently high concentrations of histidine (Tanaka and Fink 1985).] Indeed, AVY200 (*sln2 P<sub>GAL</sub>-UBR1*) cells were found to grow on SD plates containing histidine at 200  $\mu$ g/ml, but could not grow at the standard histidine concentration of 20  $\mu$ g/ml (Fig. 2 B).

To examine the histidine import of AVY200 (*sln2*) cells in a different way, we compared the uptake of L-[<sup>14</sup>C]histidine into *sln2* and other yeast strains. Since high concentrations of histidine in the medium would have interfered with the assay, we introduced the *HIS3* gene into all four of the *his3Δ* strains to be tested, making them prototrophic for histidine. As shown in Fig. 3, in the absence of Ubr1p the relative rate of histidine uptake in AVY200 (*sln2 P<sub>GAL</sub>-UBR1*) cells was about 35% of the rate in the parental JD54 (*SLN2 P<sub>GAL</sub>-UBR1*) cells. In the presence of Ubr1p (at its wildtype concentration) the rate of histidine uptake in *sln2* cells increased to about 75% of that in *SLN2* cells (Fig. 3), thus accounting, at least in part, for the observed growth of *sln2* cells on SG plates (Fig. 1 B). The incomplete recovery of histidine uptake in Ubr1p-expressing AVY200 (*sln2*) cells was consistent with their



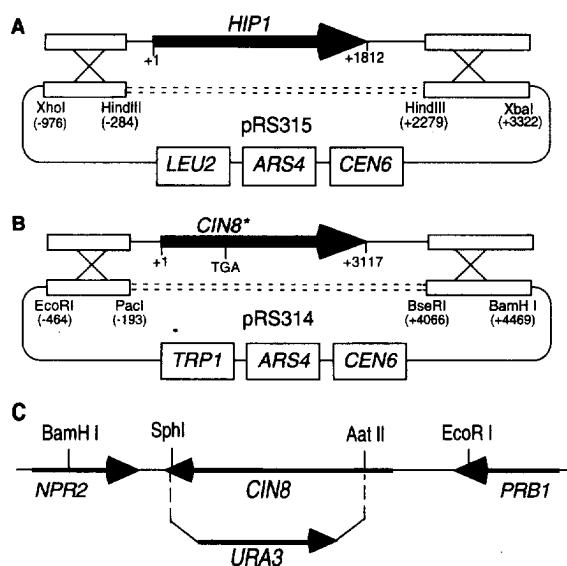
**Fig. 3** *sln2* cells are impaired in the import of histidine. *SLN2* is the parental JD54 strain (*P<sub>GAL</sub>-UBR1*) (Table 1). *hip1Δ* is a *HIP1*-lacking derivative of JD54. *sln2* (pUBR1) is the AVY200 (*sln2 P<sub>GAL</sub>-UBR1*) strain carrying pUBR1. The uptake of L-[<sup>14</sup>C]histidine was measured as described in Materials and methods. The results are expressed as the percent of uptake by *SLN2* cells, and represent the means from three independent experiments. Standard deviations are shown above the bars

detectably slower growth on SG plates in comparison to JD54 (*SLN2*) cells (Fig. 1 A,B). Note that the decreased (about 35% of normal) rate of  $^{14}\text{C}$ -histidine import in *sln2* cells was still considerably higher than the one in *hip1* $\Delta$  cells, which lack the histidine-specific transporter (Fig. 3). This observation is in agreement with the fact that the minimal concentration of histidine that allowed the growth of *sln2* cells on SD plates (75  $\mu\text{g}/\text{ml}$ ) was much lower than the concentration of histidine (200  $\mu\text{g}/\text{ml}$ ) required for the comparable growth rate of *hip1* $\Delta$  cells (data not shown).

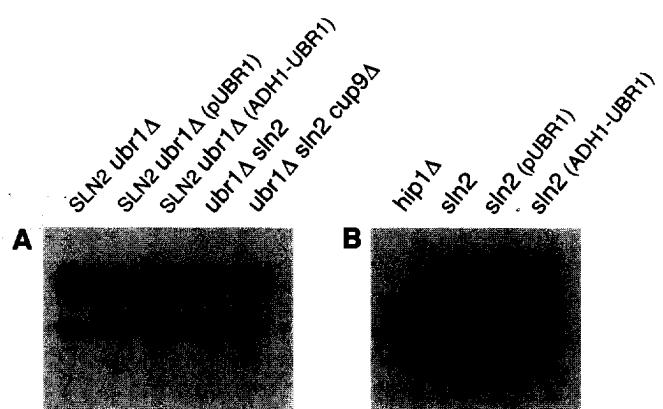
#### Down-regulation of *HIP1* in the *sln2 ubr1* $\Delta$ mutant

In *S. cerevisiae*, the import of histidine across the plasma membrane is carried out largely, if not exclusively, by two transporters (permeases), Hip1p and Gap1p (Tanaka and Fink 1985; Jauniaux and Grenson 1990). Gap1p is a general amino-acid transporter whose activity is repressed by rich nitrogen sources such as  $\text{NH}_4^+$  in SD or SG media (Jauniaux and Grenson 1990; Hein et al. 1995). Consequently, the histidine-specific permease Hip1p is the main importer of histidine under these conditions. The reduced uptake of histidine by AVY200 (*sln2*) cells in the absence of Ubr1p (in the absence of the N-end rule pathway) could be caused by a defective *HIP1* allele in *sln2* cells or by a reduced expression of the otherwise wild-type *HIP1*.

To address the first possibility, we used the method of gap repair (Fig. 4 A) (see Materials and methods) to isolate *HIP1* from both AVY200 (*sln2*) and the parental JD54 (*SLN2*) strains. Complementation testing of *HIP1* from these strains with *hip1* $\Delta$  *S. cerevisiae* showed that



**Fig. 4 A–C** Gap repair-based cloning of the *HIP1* and *CIN8\** alleles from *sln2* cells. Panels A and B illustrate, respectively, the isolation of *HIP1* and an allele of *CIN8* (*CIN8\**) from AVY200 (*sln2*  $P_{GAL}-UBR1$ ) cells. C deletion/disruption of the *CIN8* gene. See Materials and methods

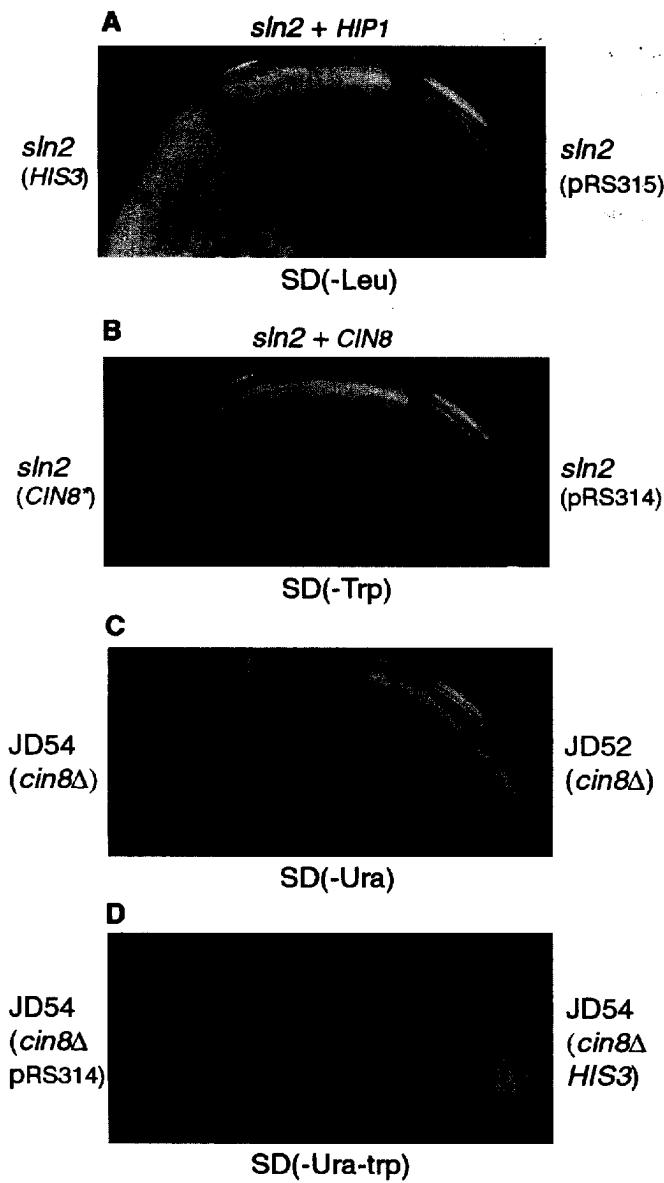


**Fig. 5** Northern-hybridization analyses of the *S. cerevisiae* *HIP1* mRNA in different genetic backgrounds. (A and B)  $^{32}\text{P}$ -labeled *HIP1* and *ACT1* (actin) DNA probes (the latter appropriately diluted) were used to determine the levels of the corresponding mRNAs in electrophoretically fractionated RNA from strains whose relevant genetic backgrounds are indicated at the top of panels. "pUBR1" and "pADH-UBR1" denote RNA from strains carrying low-copy plasmids expressing *UBR1* from either the natural *UBR1* promoter or from the  $P_{ADH1}$  promoter. See also Materials and methods

the two *HIP1* alleles were indistinguishable in their ability to restore histidine import in *hip1* $\Delta$  cells (data not shown), strongly suggesting that *sln2* cells contained the wild-type version of *HIP1*. Quantitative Northern hybridization was then used to compare the levels of *HIP1* mRNA in *sln2* and congenic *SLN2* cells in the absence and presence of Ubr1p (Fig. 5 and see Materials and methods). The relative level of *HIP1* mRNA was reproducibly about 2-fold lower in *sln2* cells lacking Ubr1p than in otherwise identical *SLN2* cells or in Ubr1p-containing *SLN2* cells (Fig. 5 A and see Materials and methods). This finding strongly suggested that the inviability of Ubr1p-lacking AVY200 (*sln2*  $P_{GAL}-UBR1$ ) cells at the standard concentration of histidine in SD medium was caused by underexpression of the Hip1p transporter. Consistent with this interpretation, we found that about a 2-fold increase in the *HIP1* copy number, through transformation with a low-copy plasmid expressing *HIP1* from its natural promoter, rescued the inviability of AVY200 cells on SD plates (Fig. 6 A). Note that the *HIS3* gene on a low-copy plasmid rescued AVY200 (*sln2*  $P_{GAL}-UBR1$ ) cells more effectively than did the plasmid-borne extra copy of *HIP1* (Fig. 6 A), again in agreement with the observed down-regulation of *HIP1* in AVY200 cells. This finding also accounted, in hindsight, for the isolation of *HIS3*, but not of *HIP1*, in our (far from exhaustive) complementation screen.

Ubr1p increases the expression of *HIP1* in *sln2* but not in *SLN2* cells

The effect of *UBR1* expression on the uptake of histidine by Ubr1p-lacking *sln2* cells (Fig. 3), and the requirement



**Fig. 6A-D** Effects of *HIP1* and *CIN8\** on the growth of strains lacking the N-end rule pathway. **A** AVY200 (*sln2 P<sub>GAL</sub>-UBRI*) cells (Table 1) carrying a low-copy, *LEU2*-marked vector *PRS315* or the otherwise identical plasmids expressing either *HIS3* or *HIP1* were grown on SD(-Leu, +0.2 mg/ml of histidine) plates to allow colony formation by all strains. Individual colonies were then streaked onto SD(-Leu) plates. Note the rescue of the AVY200 (*sln2 P<sub>GAL</sub>-UBRI*) cells by *HIP1* and their even stronger rescue by *HIS3*. **B** neither the *CIN8\** allele (isolated from *sln2* cells) nor the corresponding *TRP1*-based vector *PRS314* rescued inviability of AVY200 (*sln2 P<sub>GAL</sub>-UBRI*) cells on SD plates, in contrast to the wild-type *CIN8* gene. **C** *cin8Δ* cells (AVY202 and AVY203, respectively) derived from either the JD54 (*P<sub>GAL</sub>-UBRI*) or JD52 (*UBRI*) strains (Table 1) were streaked onto SD plates. Note inviability of *cin8Δ* cells in the absence of *UBRI* expression. Cells were grown at first on SD(-Ura, +0.2 mg/ml of histidine) plates to allow colony formation by both strains. Individual colonies were then streaked onto SD(-Ura) plates. **D** same as in C but *cin8Δ* JD54 (*cin8Δ P<sub>GAL</sub>-UBRI*) cells carried either the control vector *PRS314* or the same vector expressing *HIS3*. See Materials and methods

of Ubr1p for the viability of these cells on SD plates (Fig. 1 A-C), could be due to an increase of *HIP1* expression in the presence of Ubr1p. Comparisons of *HIP1* mRNA levels among *sln2* strains that either lacked Ubr1p or produced increasing amounts of Ubr1p showed that the expression of *HIP1* was indeed increased in the presence of Ubr1p in *sln2* cells, but not to its wild-type level (Fig. 5). As quantified with PhosphorImager and internal reference, the increase caused by the presence of Ubr1p was, reproducibly, about 1.5-fold, in comparison to the approximately 2-fold increase that would have been required for restoration of the wild-type level of *HIP1* mRNA (Fig. 5 and data not shown). The partial restoration of *HIP1* expression by Ubr1p in *sln2* cells is consistent with the observed partial recovery of histidine uptake (Fig. 3), and with lower than wild-type growth rates of Ubr1p-supplemented *sln2* cells (Fig. 1 B).

We also asked whether the presence of Ubr1p could increase the expression of *HIP1* in *SLN2* cells. As shown in Fig. 5 A, the expression of Ubr1p, at increasing levels, in *SLN2 ubr1Δ* cells did not influence the expression of *HIP1*. Taken together, our results indicated that at least two distinct pathways influence the expression of *HIP1*: a *SLN2*-dependent pathway and the *UBRI*-dependent N-end rule pathway. The latter pathway is sufficient for the viability-maintaining expression of *HIP1* in the absence of *SLN2*. Although the expression of *HIP1* was not extinguished even in the absence of both *SLN2* and *UBRI* (Fig. 5 A), it became low enough to result in growth arrest at the standard concentration of histidine in the medium (Fig. 1 A).

Previous work (Byrd et al. 1998) has shown that the N-end rule pathway strongly up-regulates the expression of *S. cerevisiae* *Ptr2p*, a peptide transporter, through the degradation of *Cup9p*, a transcriptional repressor of *PTR2*. We therefore tested whether the Ubr1p-dependent increase of *HIP1* expression in AVY200 (*sln2 P<sub>GAL</sub>-UBRI*) cells (Fig. 5) involves the degradation of *Cup9p*, which is known to either repress or activate several yeast genes (G. Turner, H. Rao, and A.V., unpublished data). A *sln2 cup9Δ P<sub>GAL</sub>-UBRI* strain was constructed; the levels of *HIP1* mRNA were compared between *sln2 CUP9* and *sln2 cup9Δ* strains, and were found to be indistinguishable (Fig. 5 A). Thus, the effect of the N-end rule pathway on the expression of *HIP1* is unlikely to involve *Cup9p*. These Northern-hybridization data (Fig. 5 A) were also consistent with the observation that a deletion of *CUP9* did not rescue the inviability of AVY200 (*sln2 P<sub>GAL</sub>-UBRI*) cells on SD plates (data not shown).

#### *SLN2* is *CIN8*

As described above, both *UBRI* and *HIS3* acted as extragenic suppressors of the *sln2* mutation. To identify the *SLN2* gene, we searched for other phenotypes of

AVY200 (*sln2 P<sub>GAL</sub>-UBR1*) cells, and found that they were inviable at 37 °C. The result of crossing AVY200 with the *SLN2* AVY199 strain (Table 1) indicated that the phenotype of temperature sensitivity (*ts*) was recessive (data not shown). Interestingly, the *ts* phenotype of *sln2* cells could not be rescued by either *UBR1* (Fig. 1 D) or *HIS3* (data not shown), indicating that the *sln2* mutation affected not only histidine uptake. Since the *sln2/SLN2* (AVY200/AVY199) diploid sporulated inefficiently, and the viability of the few spores that did form was very low, it was difficult to preclude, through tetrad analysis, the alternative possibility that *sln2* cells actually bore more than one mutant gene. Therefore we attempted to identify a gene(s) that would rescue the *ts* phenotype of *sln2* cells, and thereafter to determine whether that gene could rescue the histidine-uptake defect as well.

A standard complementation screen (see Materials and methods) yielded *CIN8*, a previously described gene encoding a kinesin-like protein (Hoyt et al. 1992; Roof et al. 1992). *CIN8* was found to alleviate not only the *ts* phenotype of *sln2* cells, but their histidine-uptake defect as well (Fig. 6 B), suggesting that the sought *SLN2* gene may be *CIN8*. To determine whether the AVY200 (*sln2 P<sub>GAL</sub>-UBR1*) strain carried a mutant allele of *CIN8*, we isolated it from these cells using the method of gap repair (Fig. 4 B and see Materials and methods). Sequencing of this allele, termed *CIN8\**, showed the presence of the A → T mutation at position 1237, resulting in a premature stop codon (TGA) at codon 413 of the 1038-codon *CIN8* ORF (data not shown). The resulting truncated product would lack the entire stalk/tail region and a portion of the putative motor domain (Hoyt et al. 1992). As expected, *CIN8\**, in contrast to *CIN8*, was unable to rescue the histidine-uptake defect (inviability on SD plates) of AVY200 cells (Fig. 6 B). These results made it very likely that the *SLN2* gene was in fact *CIN8*.

To produce independent evidence bearing on this conjecture, we deleted the bulk of *CIN8* from the JD52 (*his3Δ UBR1*) and JD54 (*his3Δ P<sub>GAL</sub>-UBR1*) strains (the latter strain is *Ubr1*<sup>-</sup> on dextrose) (see Fig. 4 C and Materials and methods). The resulting *cin8Δ* strains AVY203 (*his3Δ cin8Δ UBR1*) and AVY202 (*his3Δ cin8Δ P<sub>GAL</sub>-UBR1*) (Table 1) were tested for viability on SD plates, i.e., for complementation of the histidine uptake defect. Whereas the AVY203 (*cin8Δ UBR1*) cells were viable under these conditions, the AVY202 (*his3Δ cin8Δ P<sub>GAL</sub>-UBR1*) cells were inviable (Fig. 6 C), a set of phenotypes indistinguishable from that conferred by the *sln2* mutation. Moreover, as would have been expected if *SLN2* and *CIN8* were the same gene, the *HIS3* gene was found to rescue *Ubr1*<sup>-</sup> *cin8Δ* cells on SD plates (Fig. 6 D), identically to the effect of *HIS3* on *sln2* cells (Fig. 2 A). Taken together, these results (Figs. 4 and 6) indicated that *sln2* was a defective (truncated) allele of the *CIN8* gene.

## Discussion

We report the following results.

(1) A synthetic lethal screen for *S. cerevisiae* mutants that require the presence of the N-end rule pathway (specifically, the presence of its recognition component *Ubr1p*) for viability yielded a mutant termed *sln2*. The viability of *sln2 ubr1* cells (specifically, of *sln2 P<sub>GAL</sub>-UBR1* cells on dextrose-containing media) could be rescued by *UBR1* expressed at either normal or increased levels.

(2) The *HIS3* gene (which was absent from the parental "wild-type" strain) was found to rescue the inviability of *Ubr1p*-lacking *sln2* cells. High (>75 µg/ml) concentrations of histidine in the medium could also rescue these cells.

(3) This defect in histidine uptake, exhibited by *Ubr1p*-lacking *sln2* cells but not by *SLN2 ubr1Δ* cells, was traced to the *HIP1* gene, encoding the histidine transporter. The *HIP1* allele in *Ubr1p*-lacking *sln2* cells was wild type, but the level of *HIP1* mRNA in these cells was reproducibly about 2-fold lower than in parental (*SLN2*) cells. In addition, the relative rate of <sup>14</sup>C-histidine uptake by *Ubr1p*-lacking *sln2 HIS3* cells was about 35% of that by *SLN2 HIS3* cells and about 50% of that by *sln2 UBR1 HIS3* cells. The incomplete shutoff of histidine import in *Ubr1p*-lacking *sln2* cells was consistent with their rescue by high concentrations of histidine in the medium. Further, the incomplete restoration of histidine import in *Ubr1p*-expressing *sln2* cells was consistent with their detectably slower growth on SD plates in comparison to either *SLN2 UBR1* or *SLN2 ubr1Δ* cells. Finally, a doubling of the *HIP1* copy number in *Ubr1p*-lacking *sln2* cells partially complemented their inviability on SD plates. Thus, the histidine-uptake defect of *Ubr1p*-lacking *sln2* cells was caused at least in part by underexpression of *Hip1p*.

(4) Yet another property of the *sln2* mutant was its inviability at 37 °C, which could not be rescued by either *HIS3* or *UBR1*. This made possible the cloning of *SLN2*, which was found to be a previously characterized gene called *CIN8*, encoding a kinesin-like protein (Hoyt et al. 1992; Roof et al. 1992). The dependence of histidine transport by *cin8Δ* cells on the presence of *Ubr1p* is a previously unknown property of *cin8Δ* cells.

*Cin8p* participates in the assembly of a bipolar spindle. *cin8* mutants are viable, owing to the presence of *Kip1p*, another kinesin-related motor protein and a functional homolog of *Cin8p*; at least one of the two proteins is required for cell viability (Hoyt et al. 1992; Roof et al. 1992). Geiser et al. (1997) identified a number of other genes that are required for viability in the absence of *CIN8*. These genes encode proteins that function in cell-division pathways (spindle motors, microtubule stabilizers, checkpoint regulators).

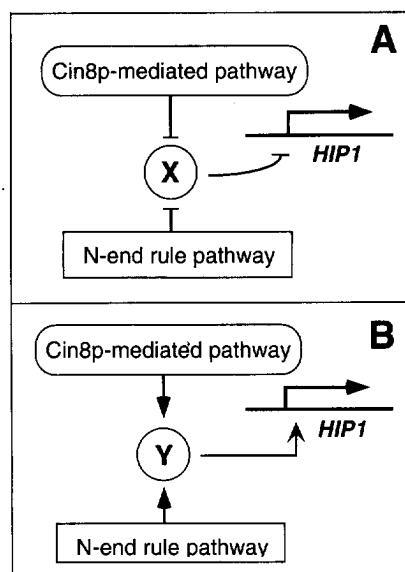
In retrospect, it was an accident of the absence of *HIS3* (rather than of another genetic marker) from the parental JD54 strain (*P<sub>GAL</sub>-UBR1*) that led us to the

finding, through a synthetic lethal screen, of a connection between a kinesin-like protein, the N-end rule pathway, and regulation of the histidine import. Preliminary tests for the ability of Ubr1p-lacking *sln2* cells to import either dipeptides or several amino acids other than histidine indicated approximately wild-type rates of uptake (data not shown).

A short-lived substrate of Ubr1p that mediates its influence on the phenotype of *cin8Δ* (*sln2*) cells is unknown. At the same time, it is clear that a connection between the N-end rule pathway and a Cin8p-dependent process involves the expression of the histidine transporter Hip1p. A parsimonious model which is consistent with the available evidence posits a short-lived transcriptional repressor of the *HIP1* gene that is degraded by the N-end rule pathway, and is also down-regulated, in an unknown manner, through a Cin8p-dependent pathway (Fig. 7 A). This model is analogous to the previously deciphered circuit that controls the expression of the peptide transporter *PTR2* in *S. cerevisiae*. It was found that Cup9p, a homeodomain-containing transcriptional repressor of *PTR2*, is a short-lived protein targeted for Ubr1p-dependent degradation, apparently through an internal degron of Cup9p (Alagramam et al. 1995; Byrd et al. 1998). In the absence of Ubr1p, the increased concentration of Cup9p results in a virtually complete shutoff of the *PTR2* gene and cessation of peptide import (Byrd et al. 1998). We know (see Results) that Cup9p is not involved in the regulation of *HIP1* expression.

An alternative model invokes a positive regulator of *HIP1* that is positively regulated by both the N-end rule and Cin8p-mediated pathways (Fig. 7 B). A transporter homolog Ssy1p of *S. cerevisiae* has been shown to be required for the expression of several amino-acid transporters, including Bap2p, Bap3p and Tat1p, and also the peptide transporter *PTR2* (Didion et al. 1998). Ssy1p appears to function at least in part as a sensor of specific amino acids in the medium. Although *SSY1* is unnecessary for histidine uptake (Didion et al. 1998), it is possible that an analogous plasma membrane-embedded sensor activates the synthesis of other transporters, including Hip1p. The absence of Cin8p may affect the secretory pathway, resulting in lower amounts of the postulated sensor in the plasma membrane. Ubr1p, the E3 component of the ubiquitin-dependent N-end rule pathway, may up-regulate this sensor; for example, through the degradation of another E3 enzyme that targets the sensor for endocytosis and degradation in the vacuole. The involvement of the ubiquitin system in the regulated endocytosis and lysosome/vacuole delivery of transmembrane proteins has previously been demonstrated in both yeast and multicellular eukaryotes (Galan et al. 1996; Hicke and Riezman 1996).

One prediction of the repressor-based model (Fig. 7 A) is that inactivation of the repressor of *HIP1* should suppress the down-regulation of *HIP1* caused by the absence of both Cin8p and Ubr1p. A genetic screen for mutants of this class is under way. An analogous earlier



**Fig. 7 A,B** Possible interconnections between the N-end rule pathway, the regulation of histidine import, and a kinesin-dependent pathway. **A** In this, apparently most-parsimonious model, a short-lived transcriptional repressor of the *HIP1* gene is degraded by the N-end rule pathway, and is also down-regulated, in an unknown manner, through a Cin8p-dependent pathway. **B** In an alternative model, a positive regulator of *HIP1* is positively regulated by both the N-end rule and Cin8p-mediated pathways

screen for suppressors of the inability of *ubr1Δ* cells to import peptides has led to the identification of Cup9p as a transcriptional repressor of the peptide transporter gene *PTR2* (Byrd et al. 1998).

**Acknowledgements** We thank members of the Varshavsky laboratory, especially Glenn Turner, Hong-Rui Wang, Fang-Yong Du, Anna Kashina, Hai Rao and Frederic Navarro-Garcia, for helpful discussions and comments on the manuscript. This work was supported by a grant to A.V. from the NIH (DK39520). Y.X. was supported by a postdoctoral fellowship from the NIH.

## References

- Alagramam K, Naider F, Becker JM (1995) A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol Microbiol* 15:225–234
- Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (eds) (1996) Current protocols in molecular biology. Wiley-Interscience, New York
- Baker RT, Varshavsky A (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J Biol Chem* 270:12065–12074
- Bachmair A, Finley D, Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234:179–186
- Bachmair A, Varshavsky A (1989) The degradation signal in a short-lived protein. *Cell* 56:1019–1032
- Balzi E, Choder M, Chen W, Varshavsky A, Goffeau A (1990) Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J Biol Chem* 265:7464–7471
- Bartel B, Wünnigen I, Varshavsky A (1990) The recognition component of the N-end rule pathway. *EMBO J* 9:3179–3189

- Baudin A, Ozier-Kalogeropoulos O, Denoel A, Lacroute A, Cullin C (1993) A simple and efficient method for direct gene-deletion in *Saccharomyces cerevisiae*. Nucleic Acids Res 21:3329–3330
- Baumeister W, Walz J, Zühl F, Seemüller E (1998) The proteasome: paradigm of a self-compartmentalizing protease. Cell 92:367–380
- Byrd C, Turner GC, Varshavsky A (1998) The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. EMBO J 17:269–277
- Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243:1576–1583
- Chen D-C, Yang B-C, Kuo T-T (1992) One-step transformation of yeast in stationary phase. Curr Genet 21:83–84
- Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 65:801–817
- deGroot RJ, Rümenapf T, Kuhn RJ, Strauss JH (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc Natl Acad Sci USA 88:8967–8971
- Didion T, Regenberg B, Jørgesen MU, Kielland-Brandt MC, Andersen HA (1998) The permease homologue Ssy1p controls the expression of amino-acid and peptide transporter genes in *Saccharomyces cerevisiae*. Mol Microbiol 27:643–650
- Dohmen RJ, Wu P, Varshavsky A (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science 263:1273–1276
- Galan JM, Moreau V, Andre B, Volland C, Haguenuuer-Tsapis R (1996) Ubiquitination mediated by the Npl1p/Rsp5b ubiquitin-protein ligase is required for endocytosis of the yeast uracil permease. J Biol Chem 271:10946–10952
- Geiser JR, Schott TJ, Kingsbury TJ, Cole NB, Totis LJ, Bhattacharyya G, He L, Hoyt MA (1997) *Saccharomyces cerevisiae* genes required in the absence of the *CIN8*-encoded spindle motor act in functionally diverse mitotic pathways. Mol Cell Biol 8:1035–1050
- Ghislain M, Dohmen RJ, Levy F, Varshavsky A (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. EMBO J 15:4884–4899
- Glickman MH, Rubin DM, Fried VA, Finley D (1998) The regulatory particle of the *Saccharomyces cerevisiae* proteasome. Mol Cell Biol 18:3149–3162
- Gonda DK, Bachmair A, Wünning I, Tobias JW, Lane WS, Varshavsky A (1989) Universality and structure of the N-end rule. J Biol Chem 264:16700–16712
- Gonen H, Schwartz AL, Ciechanover A (1991) Purification and characterization of a novel protein that is required for degradation of N-alpha-acetylated proteins by the ubiquitin system. J Biol Chem 266:19221–19231
- Grauslund M, Didion T, Kielland-Brandt MC, Andersen HA (1995) *BAP2*, a gene encoding a permease for branched-chain amino acids in *Saccharomyces cerevisiae*. Biochim Biophys Acta 1269:275–280
- Hein C, Springael J-Y, Christiane V, Haguenuuer-Tsapis R, Andre B (1995) *NPI1*, an essential yeast gene involved in induced degradation of Gap1 and Fur4 permeases, encodes the Ras5 ubiquitin-protein ligase. Mol Microbiol 18:77–87
- Hershko A (1991) The ubiquitin pathway for protein degradation. Trends Biochem Sci 16:265–268
- Hershko A (1997) Roles of ubiquitin-mediated proteolysis in cell cycle control. Curr Opin Cell Biol 9:788–799
- Herskowitz I, Jensen RE (1991) Putting the *HO* gene to work: practical uses for mating-type switching. Methods Enzymol 194:132–146
- Hicke L, Riezman H (1996) Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. Cell 84:277–287
- Hill CP, Johnston NL, Cohen RE (1993) Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. Proc Natl Acad Sci USA 90:4136–4140
- Ho AK, Racznak GA, Ives EB, Wente SR (1998) The integral membrane protein snl1p is genetically linked to yeast nuclear pore complex function. Mol Biol Cell 9:355–373
- Hochstrasser M (1996) Ubiquitin-dependent protein degradation. Annu Rev Genet 30:405–439
- Hondermarck H, Sy J, Bradshaw RA, Arfin SM (1992) Di peptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. Biochem Biophys Res Commun 30:280–288
- Hoyt MA, He L, Loo KK, Saunders WS (1992) Two *Saccharomyces cerevisiae* kinesin-related gene products required for mitotic spindle assembly. J Cell Biol 118:109–120
- Jauniaux J-C, Grenson M (1990) *GAPI*, the general amino-acid permease gene of *Saccharomyces cerevisiae*: nucleotide sequence, protein similarity with the other bakers yeast amino-acid permeases, and nitrogen catabolite repression. Eur J Biochem 190:39–44
- Johnson ES, Ma PC, Ota IM, Varshavsky A (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. J Biol Chem 270:17442–17456
- Kwon YT, Reiss Y, Fried VA, Hershko A, Yoon JK, Gonda DK, Sangan P, Copeland NG, Jenkins NA, Varshavsky A (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc Natl Acad Sci USA 95:7898–7903
- Madura K, Varshavsky A (1994) Degradation of G-alpha by the N-end rule pathway. Science 265:1454–1458
- Merrill BJ, Holm C (1998) The RAD52 recombinational repair pathway is essential in pol30 (PCNA) mutants that accumulate small single-stranded DNA fragments during DNA synthesis. Genetics 148:611–624
- Mumberg D, Muller R, Funk M (1994) Regulatable promoters of *Saccharomyces cerevisiae* – comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res 22:5767–5768
- Ota IM, Varshavsky A (1993) A yeast protein similar to bacterial two-component regulators. Science 262:566–569
- Peters J-M, King RW, Deshaies RJ (1998) Cell cycle control by ubiquitin-dependent proteolysis. Ubiquitin and the biology of the cell. Plenum Press, New York, pp 345–387
- Pickart CM (1997) Targeting of substrates to the 26S proteasome. FASEB J 11:1055–1066
- Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H (1996) Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 “two-component” osmosensor. Cell 86:865–875
- Rechsteiner M, Hoffman L, Dubiel W (1993) The multicatalytic and 26S proteases. J Biol Chem 268:6065–6068
- Roof DM, Meluh PB, Rose MD (1992) Kinesin-related proteins required for assembly of the mitotic spindle. J Cell Biol 118:95–108
- Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol 194:281–301
- Scheffner M, Smith S, Jentsch S (1998) The ubiquitin conjugation system. In: Peters J-M, Harris JR, Finley D (eds) Ubiquitin and the biology of the cell. Plenum Press, New York, pp 65–98
- Schmitt ME, Brown TA, Trumppower BL (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res 18:3091–3092
- Sherman F, Fink GR, Hicks JB (1986) Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Sijts AJ, Pilip I, Pamer EG (1997) The *Listeria* monocytogenes-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class-I antigen processing of bacterial proteins. J Biol Chem 272:19261–19268
- Sikorski RS, Hieter P (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. Genetics 122:19–27

## Analysis of a conditional degradation signal in yeast and mammalian cells

Frédéric Lévy\*, Jennifer A. Johnston† and Alexander Varshavsky

Division of Biology, California Institute of Technology, Pasadena, CA, USA

The N-end rule pathway is a ubiquitin-dependent proteolytic system, the targets of which include proteins that bear destabilizing N-terminal residues. The latter are a part of the degradation signal called the N-degron. Arg-DHFR<sup>ts</sup>, an engineered N-end rule substrate, bears N-terminal arginine (a destabilizing residue) and DHFR<sup>ts</sup> [a temperature-sensitive mouse dihydrofolate reductase (DHFR) moiety]. Previous work has shown that Arg-DHFR<sup>ts</sup> is long-lived at 23 °C but short-lived at 37 °C in the yeast *Saccharomyces cerevisiae*. In the present work, we extended this analysis, and found that the degradation of Arg-DHFR<sup>ts</sup> can be nearly completely inhibited *in vivo* by methotrexate (MTX), a low-*M<sub>r</sub>* ligand of DHFR. In *S. cerevisiae*, Arg-DHFR<sup>ts</sup> is degraded at 37 °C exclusively by the N-end rule pathway, whereas in mouse cells the same protein at the same temperature is also targeted by another proteolytic system, through a degron in the conformationally perturbed DHFR<sup>ts</sup> moiety. In mouse cells, MTX completely inhibits the degradation of Arg-DHFR<sup>ts</sup> through its degron within the DHFR<sup>ts</sup> moiety, but only partially inhibits degradation through the N-degron. When the N-terminus of Arg-DHFR<sup>ts</sup> was extended with a 42-residue lysine-lacking extension, termed e<sup>ΔK</sup>, the resulting Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup> was rapidly degraded at both 23 °C and 37 °C. Moreover, the degradation of Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>, in contrast with that of Arg-DHFR<sup>ts</sup>, could not be inhibited by MTX, suggesting that the metabolic stability of Arg-DHFR<sup>ts</sup> at 23 °C results, at least in part, from steric inaccessibility of its N-terminal arginine. The N-degron of Arg-DHFR<sup>ts</sup> is the first example of a portable degradation signal the activity of which can be modulated *in vivo* by a cell-penetrating compound. We discuss implications of this advance and the mechanics of targeting by the ubiquitin system.

**Keywords:** degron; methotrexate; N-end rule; proteolysis; ubiquitin.

A number of regulatory circuits, including those that control the cell cycle, cell differentiation and responses to stress, involve metabolically unstable proteins [1–6]. A short *in vivo* half-life of a regulator provides a way to generate its spatial gradient and allows rapid adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis. A protein can also be conditionally unstable, i.e. long-lived or short-lived, depending on the state of the cell. One example of the latter class are cyclins, a family of related proteins the destruction of which at specific stages of the cell cycle regulates cell division and growth [3,7].

Features of proteins that confer metabolic instability are called degradation signals, or degrons [8]. An essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein [9,10]. A set of N-degrons containing different N-terminal residues which are destabilizing in a given cell yields a rule, termed the N-end rule, which relates

the *in vivo* half-life of a protein to the identity of its N-terminal residue. In eukaryotes, the N-end rule pathway is a part of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, differentiation and responses to stress [2,4,11–13]. In many of these settings, Ub acts through routes that involve the degradation of Ub–protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease [14,15].

In eukaryotes, linear Ub–protein fusions are rapidly cleaved by Ub-specific proteases (UBPs) at the Ub–protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues, a technical advance that led to the discovery of the N-end rule [9,10]. When mouse dihydrofolate reductase (DHFR), a 20-kDa monomeric protein, was expressed as a Ub-Arg-DHFR fusion in the yeast *Saccharomyces cerevisiae* at 30 °C, the resulting Arg-DHFR (produced through deubiquitylation) was long-lived ( $t_{0.5} > 4$  h), even though it bore N-terminal Arg, a destabilizing residue [16]. By contrast, a modified protein, Arg-e<sup>ΔK</sup>-DHFR (produced from Ub-Arg-e<sup>ΔK</sup>-DHFR), which bore a 42-residue, lysine-containing extension (denoted as e<sup>ΔK</sup>) between the N-terminal Arg residue and the first (Val) residue of DHFR, was short-lived *in vivo* ( $t_{0.5} \approx 10$  min), being degraded by the N-end rule pathway [16].

This and other evidence indicated that the N-degron comprises two essential determinants: a destabilizing N-terminal residue and an internal Lys residue (or residues) of a substrate. The Lys residue is the site of attachment of a multi-Ub chain [16–18]. Arg-e<sup>ΔK</sup>-DHFR contains a complete N-degron because it bears both an N-terminal destabilizing residue (Arg) and, in the e<sup>ΔK</sup> extension, at least one targetable Lys residue. In contrast, the N-degron of Arg-DHFR is incomplete, for at least one of the

\*Correspondence to Alexander Varshavsky, Division of Biology, 147–75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125, USA.  
Tel.: +1-626-395-3785, Fax: +1-626-440-9821.

E-mail: avarsh@cco.caltech.edu

†Present address: Ludwig Institute for Cancer Research, Lausanne Branch, Ch. des Boveresses 155, CH-1066 Epalinges, Switzerland.

‡Present address: Department of Biological Sciences, Stanford University, CA 94305, USA.

Abbreviations: βgal, *Escherichia coli* β-galactosidase; DHFR, mouse dihydrofolate reductase; ID, initial decay; MTX, methotrexate; Ub, ubiquitin; UBP, Ub-specific protease; UPR, ubiquitinproteinreference.

Note: F. Lévy and J. A. Johnston contributed equally to the present work.

(Received 28 July 1998; revised 12 October 1998; accepted 13 October 1998)

two mutually non-exclusive reasons: the absence of Lys residues accessible to the targeting complex of the N-end rule pathway and/or poor accessibility of the N-terminal Arg in Arg-DHFR to the same targeting complex. Since the 20-kDa mouse DHFR contains 16 Lys residues, this interpretation presumes that, in a folded DHFR molecule, the lysines are ineffective as ubiquitylation sites, because of the lysine's lack of mobility and/or its distance from a destabilizing N-terminal residue [10].

Previous work [19] described a temperature-sensitive (*ts*) allele of DHFR that was used to construct a heat-inducible N-degron, in a Ub-Arg-DHFR<sup>ts</sup>-Ura3p fusion which contained *S. cerevisiae* Ura3p as a C-terminal reporter. The Ub-Arg-DHFR<sup>ts</sup> moiety of this fusion was identical with Ub-Arg-DHFR above, except for a Pro → Leu alteration at position 66 of the DHFR moiety. Arg-DHFR<sup>ts</sup>-Ura3p was long-lived at 23 °C but short-lived at 37 °C in yeast. Moreover, the Ub-Arg-DHFR<sup>ts</sup> module of Ub-Arg-DHFR<sup>ts</sup>-Ura3p was shown to be portable, in that linking it to a protein of interest conferred heat-inducible metabolic instability on the entire fusion [19].

In addition to yielding a new method for the construction of *ts* mutants, the heat-inducible N-degron [19] provided an approach to analyzing the mechanism of targeting by the N-end rule pathway. In carrying out this analysis, described below, we have found that the degradation of Arg-DHFR<sup>ts</sup> can be inhibited *in vivo* by methotrexate (MTX), a low-*M<sub>r</sub>* ligand of DHFR. We also show that, in mouse cells, Arg-DHFR<sup>ts</sup> is targeted by both the N-end rule pathway and another proteolytic system, whereas in yeast the same test protein is exclusively an N-end rule substrate. The N-degron of Arg-DHFR<sup>ts</sup> is the first example of a

portable degradation signal whose activity can be modulated *in vivo* by a cell-penetrating compound.

## MATERIALS AND METHODS

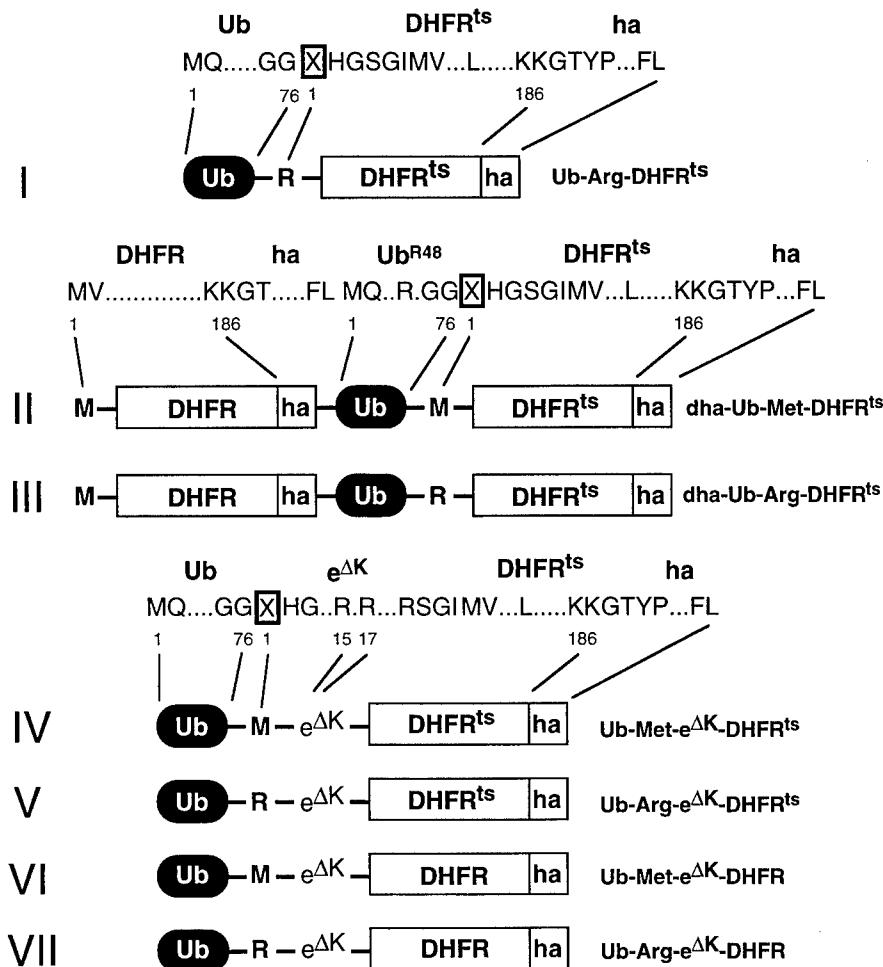
### Plasmids for expression in *S. cerevisiae*

The plasmid pJJRtd, which expressed Ub-Arg-DHFR<sup>ts</sup>-ha (Fig. 1, construct I), was based on the vector pRS316 [20]. A 1.0-kb *Eco*RI/*Hind*III fragment of pPW58 [8] was ligated into pRS316, followed by the cloning of the *P<sub>CUP1</sub>* promoter-containing 0.4 kb *Eco*RI/*Sac*I fragment of pPW48 [19] into the *Eco*RI site, a step that yielded pJJRtd. pJJMtd, which expressed Ub-Met-DHFR<sup>ts</sup>-ha, was prepared by ligating the large *Age*I/*Af*II fragment of pJJRtd to a small fragment from the *Age*I/*Af*II digest of pLGUB-M-DHFRha [21].

### Plasmids for expression in mammalian cells

Mouse DHFR and DHFR<sup>ts</sup> open reading frames bearing the desired restriction sites at their 5' and 3' ends were produced using PCR. All DHFR and DHFR<sup>ts</sup> moieties were flanked, at their C-termini, with the ha epitope [22] (Fig. 1), and were cloned into the pRc/CMV vector (Invitrogen, San Diego, CA, USA).

Constructs II and III (Fig. 1) were produced by PCR amplification of a fragment encoding Met-DHFR<sup>ts</sup> and Arg-DHFR<sup>ts</sup>, respectively, using construct I as a template. The N-terminal Met of construct II was specified by the PCR primer



**Fig. 1. Test proteins.** Fusions used in this work contained some of the following moieties: the Ub moiety, either N-terminal (constructs I, IV–VII) or placed between other moieties (constructs II and III); an Arg residue at the junction between Ub and the C-terminal part of the fusion (constructs I, III, V and VII); a Met residue at the same junction (constructs II, IV, and VI); a 42-residue *E. coli* Lac repressor-derived sequence, termed e<sup>ΔK</sup> [extension (e) lacking lysines (ΔK)], between Ub and the reporter part of the fusion (see the main text and [22,27]) (constructs IV–VII); DHFRha, a mouse DHFR moiety extended at the C-terminus by a sequence containing the hemagglutinin-derived ha epitope (constructs I–VII); DHFR<sup>ts</sup>, a mutant temperature-sensitive DHFR moiety that differs from DHFR by the Pro → Leu alteration at position 66 [19] (see the main text). Amino acid residues are indicated by their single-letter abbreviations. Residues that vary between constructs are boxed in the sequences above the diagrams.

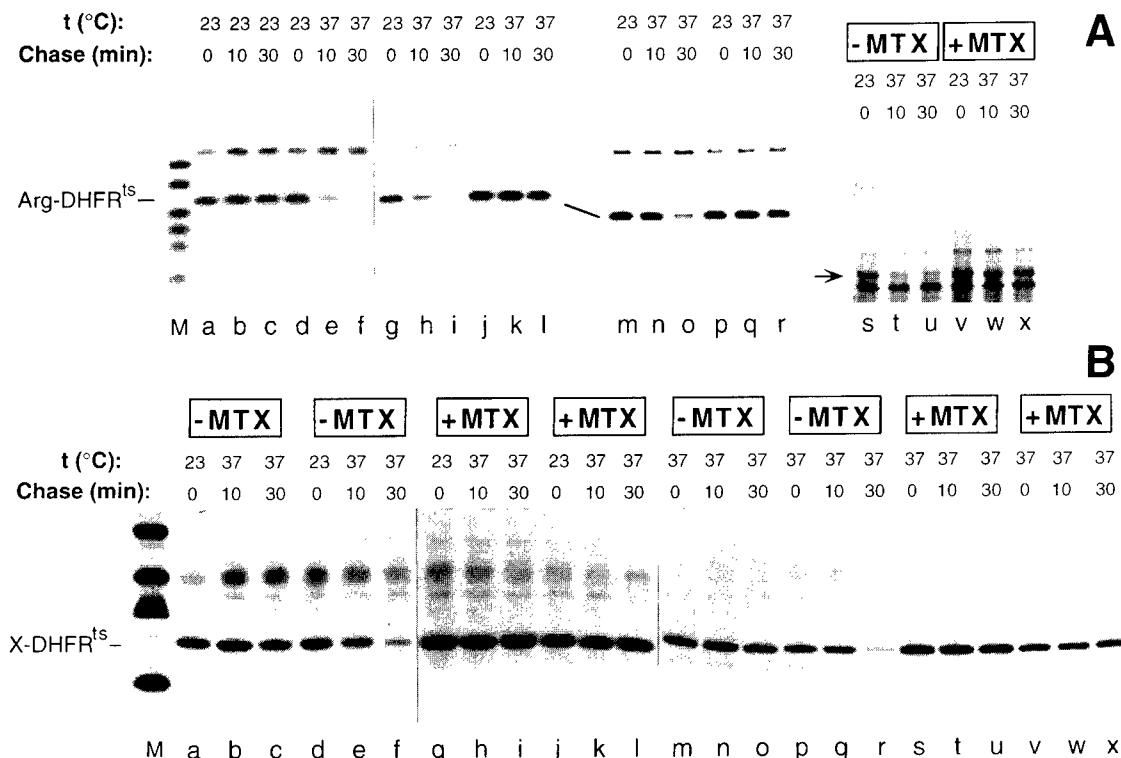
annealing to the 5' end of the fragment. The PCR products were digested with *Xba*I, blunted with Klenow Pol I, followed by a cut with *Sac*II. A fragment bearing an open reading frame was inserted between the blunt-ended *Not*I site and the *Sac*II site of the plasmid pRc/dhaUbM<sub>β</sub>gal [23], resulting in the plasmids pRc/dha-Ub-Met-DHFR<sup>ts</sup> and pRc/dha-Ub-Arg-DHFR<sup>ts</sup>. Met- and Arg-DHFR<sup>ts</sup> contained the six-residue sequence His-Gly-Ser-Gly-Ile-Met between the N-terminal residue and Val, the first natural residue of DHFR (Fig. 1).

To produce constructs IV–VII (Fig. 1), a fragment encoding Arg-DHFR<sup>ts</sup> was amplified by PCR, using construct I as template. The PCR product was digested with *Xba*I, and blunted with Klenow Pol I, followed by a *Sac*II cut. The resulting fragment was inserted between the blunt-ended *Not*I site and the *Sac*II site of pRc/UbS<sub>β</sub>gal, yielding pRc/Ub-Arg-DHFR<sup>ts</sup>. The latter served as a vector for the constructs IV–VII. Constructs IV and V were obtained by amplifying fragments encoding, respectively, Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup> and Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>, using the plasmid pLG/Ub<sup>V76</sup>-Val-e<sup>ΔK</sup>DHFRha [22] as a template. The N-terminal residues Met and Arg were specified by a PCR primer annealing to the 5' end of the fragment. Digestion of the PCR products with *Sac*II and *PflMI* yielded a fragment coding

for Met- or Arg-e<sup>ΔK</sup> and the first 39 residues of DHFR (DHFR<sub>1–39</sub>). (*PflMI* cleaves the DHFR<sup>ts</sup> open reading frame upstream of the mutated codon at position 66.) This fragment was inserted in-frame between the *Sac*II/*PflMI* sites of the plasmid pRc/Ub-Arg-DHFR<sup>ts</sup>, yielding plasmids expressing Ub-Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup> and Ub-Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>. Constructs VI and VII were produced by exchanging a *SnaBI/PflMI* fragment of pRc/dhaUbNe<sup>K</sup>DHFRha with the *SnaBI/PflMI* fragments derived from constructs IV and V, respectively. This step eliminated the fragment encoding dha-Ub-Asn-e<sup>K</sup>-DHFR<sub>1–13</sub> of the original plasmid, replacing it, in-frame, with the sequence encoding either Ub-Met-e<sup>ΔK</sup>-DHFR<sub>1–13</sub> or Ub-Arg-e<sup>ΔK</sup>-DHFR<sub>1–13</sub>. (The plasmid pRc/dhaUbNe<sup>K</sup>DHFRha was constructed by replacing the region encoding βgal in pRc/dhaUbX<sub>β</sub>gal [23] with the region encoding DHFR.)

#### Yeast and mouse cell cultures, transfection, and pulse-chase analysis

*S. cerevisiae* were grown and manipulated as described previously [24] and in the legend to Fig. 2. Mouse L cells, a fibroblast-like cell line (ATCC CCL 1.3, American Type



**Fig. 2. Inhibition of Arg-DHFR<sup>ts</sup> degradation by MTX in *S. cerevisiae*.** (A) Lane M, molecular mass markers. Lanes a–c, JD47-13c (*UBR1*) cells expressing Arg-DHFR<sup>ts</sup> (Ub-Arg-DHFR<sup>ts</sup>) were labeled for 4 min at 23 °C with [<sup>35</sup>S]methionine/cysteine, followed by a chase for 10 and 30 min as described [40], then extraction, immunoprecipitation, and Tricine-based SDS/PAGE (12% gel) of Arg-DHFR<sup>ts</sup> [41]. Lanes d–f, same as lanes a–c, except that immediately after the 4-min pulse cells were shifted to 37 °C, and the chase was carried out at 37 °C. Lanes g–i, same as lanes d–f, but an independent pulse–chase experiment. Lanes j–l, same as lanes g–i, but with JD55 (*ubr1Δ*) *S. cerevisiae*. Lanes m–o, same as lanes g–i, but with KMY613 (*ubc2Δ*) *S. cerevisiae* strain expressing the plasmid-borne *UBC2* gene [42]. Lanes p–r, same as lanes m–o, but with KMY613 (*ubc2Δ*) lacking the plasmid-borne *UBC2*. Lanes s–u, JD47-13c (*UBR1*) cells expressing Arg-DHFR<sup>ts</sup>-ha-Cdc28p (Ub-Arg-DHFR<sup>ts</sup>-ha-Cdc28p) [19] were labeled and processed as described for lanes d–f. Lanes v–x, same as lanes s–u, but with MTX added to the growth medium to a final concentration of 20 μM 30 min before the pulse. (B) Lane M, molecular mass markers. Lanes a–c, JD47-13c (*UBR1*) cells expressing Met-DHFR<sup>ts</sup> (Ub-Met-DHFR<sup>ts</sup>) were labeled for 4 min at 23 °C with [<sup>35</sup>S]methionine/cysteine. Immediately thereafter, the cells were shifted to 37 °C, followed by a chase at 37 °C for 10 and 30 min, extraction, immunoprecipitation, and SDS/PAGE of Met-DHFR<sup>ts</sup>. Lanes d–f, same as lanes a–c, but with Arg-DHFR<sup>ts</sup>. Lanes g–i, same as lanes a–c, but with MTX added to the growth medium 30 min before the pulse to a final concentration of 20 μM. Lanes j–l, same as lanes g–i, but with Arg-DHFR<sup>ts</sup>. Lanes m–o, same as lanes a–c, except that both the pulse and the chase were at 37 °C. Lanes p–r, same as lanes m–o, but with Arg-DHFR<sup>ts</sup>. Lanes s–u, same as lanes m–o, but with MTX added to the growth medium 30 min before the pulse to a final concentration of 20 μM. Lanes v–x, same as lanes s–u, but with Arg-DHFR<sup>ts</sup>.

Culture Collection, Rockville, MD, USA), were grown as monolayers in Dulbecco's modified Eagle's medium/F12 medium, supplemented with 10% fetal bovine serum, antibiotics, 2 mM L-glutamine and 20 mM Hepes (sodium salt; pH 7.3). The cultures were regularly checked for the absence of mycoplasmas. Transient transfections and pulse-chase analyses were performed as described [23]. For the pulse-chases at 23 °C, cells were transferred to this temperature 30 min before labeling for 10 min with 0.1 mCi (1 Ci = 37 GBq) of [<sup>35</sup>S]EXPRESS (New England Nuclear, Boston, MA, USA). Cells were chased at either 23 °C or 37 °C in the complete medium containing 0.1 mg·mL<sup>-1</sup> cycloheximide. All media were pre-equilibrated at the desired temperature. In experiments that involved MTX, the drug was added 30 min before labeling and was present at a final concentration of 20 μM (diluted from 20 mM MTX in 0.9 M potassium/sodium phosphate, pH 7.3). The samples were analyzed by SDS/PAGE (15% gel), followed by autoradiography and quantitation using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA).

To distinguish between partial half-lives that approximate the slopes of different regions of a non-exponential decay curve, a generalized half-life term,  $t_{0.5}^{y-z}$ , was used, where 0.5 denotes the parameter's half-life aspect and  $y-z$  denotes the relevant time interval, from  $y$  to  $z$  min of chase. Another term [23], initial decay (ID), expressed as a percentage, equals 1 minus the ratio of the amount of a radiolabeled X-DHFR test protein (normalized against the reference protein, Met-DHFR-Ub; Fig. 1) at the end of the pulse (time 0) to the (normalized) amount of a radiolabeled long-lived X-DHFR such as Met-DHFR. As ID may depend on the duration of a pulse, a superscript, in ID<sup>x</sup>, invokes the pulse time explicitly [23].

## RESULTS

### Inhibition of Arg-DHFR<sup>ts</sup> degradation by MTX in *S. cerevisiae*

The major test protein of this work was Arg-DHFR<sup>ts</sup> (produced from Ub-Arg-DHFR<sup>ts</sup>) (Fig. 1, construct I), identical with the Arg-DHFR<sup>ts</sup> moiety of a larger fusion described in the introduction. The test proteins bore C-terminal *ha*, a hemagglutinin-derived epitope tag recognized by a monoclonal antibody [22] (Fig. 1; the *ha* epitope is omitted below in the names of specific test proteins). For the analysis in *S. cerevisiae*, Arg-DHFR<sup>ts</sup> was expressed from the copper-inducible P<sub>CUP1</sub> promoter in a low-copy vector, and the metabolic fate of Arg-DHFR<sup>ts</sup> was determined in a pulse-chase assay (see Materials and methods). Similarly to the previously examined Arg-DHFR<sup>ts</sup>-Ura3p and Arg-DHFR<sup>ts</sup>-Cdc28p [19], Arg-DHFR<sup>ts</sup> was a long-lived protein at 23 °C ( $t_{0.5} > 4$  h) but a short-lived protein at 37 °C ( $t_{0.5} \approx 4$  min) (Fig. 2A, lanes a-c vs. d-f). The degradation of Arg-DHFR<sup>ts</sup> at 37 °C was carried out by the N-end rule pathway, as indicated by (a) the requirement for Ubr1p (E3), the recognition component of this pathway (Fig. 2A, lanes g-i vs. j-l), (b) the requirement for Ubc2p (E2), the relevant Ub-conjugating enzyme (Fig. 2A, lanes m-o vs. p-r), and (c) the long half-life of Met-DHFR<sup>ts</sup>, an otherwise identical protein that bore N-terminal Met, a stabilizing residue, instead of Arg (Fig. 2B, lanes a-c vs. d-f).

Previous work [27] has shown that the degradation of Arg-e<sup>K</sup>-DHFR by the N-end rule pathway in an extract from rabbit reticulocytes can be inhibited by the folate analog MTX, a high-affinity DHFR ligand ( $K_d \approx 10$  pm). It was far from clear whether MTX would have the same effect *in vivo*, in part because the purified N-end rule substrates added to the extract

contained the folded DHFR moieties [27]. By contrast, a nascent DHFR-based substrate is unable to bind MTX until after the folding of DHFR [28], but can be targeted by the N-end rule pathway *in vivo* at any time, possibly even during translation [16].

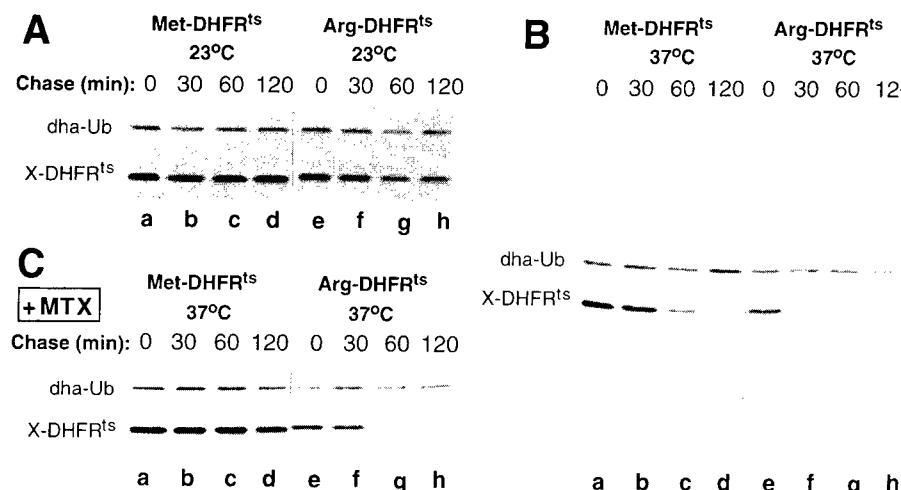
The metabolic fate of Arg-DHFR<sup>ts</sup> at 37 °C was monitored in *S. cerevisiae* in the absence or presence of 20 μM MTX in the medium. MTX was found to be an efficacious inhibitor of Arg-DHFR<sup>ts</sup> degradation in *S. cerevisiae* (Fig. 2B, lanes d-f and p-r vs. j-l and v-x). MTX inhibited the degradation of Arg-DHFR<sup>ts</sup> regardless of whether it was added to the growth medium 15 min or 24 h before the pulse-chase assay (data not shown; the growth of cells was not affected significantly by 20 μM MTX for at least 24 h). The protective effect of MTX on Arg-DHFR<sup>ts</sup> was essentially the same regardless of whether the pulse labeling (followed by the 37 °C chase) was carried out at 23 °C (Fig. 2B, lanes d-f vs. j-l) or at 37 °C (Fig. 2B, lanes p-r vs. v-x).

The degradation (at 37 °C) of Arg-DHFR<sup>ts</sup>-Cdc28p, which differed from Arg-DHFR<sup>ts</sup> (produced from Ub-Arg-DHFR<sup>ts</sup>) by the presence of the Cdc28p kinase moiety, was also inhibited by MTX (Fig. 2A, lanes s-u vs. v-x). Arg-DHFR<sup>ts</sup>-Cdc28p, which was long-lived at 23 °C, or at 37 °C in *ubr1Δ* cells, could function as a Cdc28p kinase *in vivo* [19]. As Arg-DHFR<sup>ts</sup>-Cdc28p was short-lived at 37 °C in *UBR1* *S. cerevisiae* (Fig. 2A, lanes s-u), the cells that also carried a mutant (*ts*) *cdc28* allele [19] were not viable at 37 °C, owing to the absence of the essential Cdc28p kinase at this temperature. The addition of MTX (at 20 μM) was found to rescue these cells at 37 °C through the inhibition of degradation of Arg-DHFR<sup>ts</sup>-Cdc28p (Fig. 2A, lanes s-u vs. v-x, and data not shown). The effect of MTX was specific for DHFR-containing substrates, as degradation of unrelated test proteins, for example Arg-βgal (Ub-Arg-βgal), by the N-end rule pathway was unimpaired in the presence of MTX (data not shown).

### DHFR as an N-end rule substrate in mammalian cells

DHFR-based substrates of the N-end rule pathway were used in the earlier studies with *S. cerevisiae* [16,19], but have not been examined in mammalian cells. We used L cells, a mouse fibroblast-like cell line, and the ubiquitin/protein/reference (UPR) technique, which increases the accuracy of pulse-chase assays [26] by providing a reference protein [23]. In the UPR technique, Ub is located between a protein of interest and a reference protein in a linear fusion (e.g. construct II in Fig. 1). This fusion is cleaved, cotranslationally or nearly so, by UBP<sub>s</sub> at the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing a C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse-chase assay, the relative amounts of the protein of interest can be normalized against the reference in the same sample. The UPR technique can thus compensate for the scatter of immunoprecipitation yields, sample volumes and other sources of sample-to-sample variation [23].

The UPR constructs of the present work were fusions containing the metabolically stable Arg-e<sup>K</sup>-DHFR-Ub moiety as a reference protein (termed dha-1) and a DHFR-based N-end rule substrate such as, for example, dha-2, as a protein of interest (Fig. 1, construct III). To prevent the possibility that the C-terminal Ub moiety of dha-1 functioned as a ubiquitylation/degradation signal, the Lys<sup>1</sup> residue of Ub (a major site of isopeptide bonds in multi-Ub chains [17,22,29]) was converted into Arg, which cannot be ubiquitylated [12].



**Fig. 3. Inhibition of Arg-DHFR<sup>ts</sup> degradation in mouse cells by MTX.** (A) Lanes a–d, mouse L cells transiently expressing a UPR-based version of Met-DHFR<sup>ts</sup> (Met-DHFR-ha-Ub<sup>R48</sup>-Met-DHFR<sup>ts</sup>-ha; see Fig. 1, construct II) were labeled at 23 °C for 10 min with [<sup>35</sup>S]methionine/cysteine, followed by a chase, also at 23 °C, for 30, 60 and 120 min in the presence of cycloheximide, extraction, immunoprecipitation, and SDS/PAGE of Met-DHFR<sup>ts</sup>-ha and Met-DHFR-ha-Ub<sup>R48</sup>. Lanes e–h, same as lanes a–d but with Arg-DHFR<sup>ts</sup> (Met-DHFR-ha-Ub<sup>R48</sup>-Arg-DHFR<sup>ts</sup>-ha). (B) Lanes a–d, same as lanes a–d in (A), except that both the pulse and the chase were at 37 °C. Lanes e–h, same as lanes a–d, but with Arg-DHFR<sup>ts</sup>-ha (Met-DHFR-ha-Ub<sup>R48</sup>-Arg-DHFR<sup>ts</sup>-ha). (C) Lanes a–d, same as lanes a–d in (B) but with 20 μM MTX in the medium. Lanes e–h, same as lanes a–d, but with Arg-DHFR<sup>ts</sup> (Met-DHFR-ha-Ub<sup>R48</sup>-Arg-DHFR<sup>ts</sup>-ha). The bands of Arg/Met-DHFR<sup>ts</sup>-ha and Met-DHFR-ha-Ub<sup>R48</sup> (denoted as dha-Ub) are indicated. A larger area of the gel is shown for lanes a–h in (B), to illustrate the overall immunoprecipitation pattern.

These fusions were expressed from the cytomegalovirus early promoter, P<sub>CMV</sub>, in mouse L cells. The cells were labeled for 10 min at either 23 °C or 37 °C with [<sup>35</sup>S]methionine/cysteine, followed by a chase of 30, 60 and 120 min at the same temperature in the presence of cycloheximide, immunoprecipitation with anti-ha monoclonal antibody, and analysis/quantitation of immunoprecipitated proteins by SDS/PAGE and Phosphor-Imager, using UPR (see Materials and methods). Pulse-chase assays in which cycloheximide (a translation inhibitor) was omitted yielded similar results (data not shown).

Both Met-DHFR<sup>ts</sup> and Arg-DHFR<sup>ts</sup> were long-lived in mouse cells at 23 °C ( $t_{0.5} > 10$  h) (Fig. 3A, lanes a–h, and Fig. 4A). By contrast, Arg-DHFR<sup>ts</sup> was short-lived at 37 °C in mouse cells: its half-life, determined between 0 and 60 min of chase and denoted as  $t_{0.5}^{0-60}$  [23], was ≈ 10 min (Fig. 3B, lanes e–h, and Fig. 4A). In addition, a large fraction (≈ 30%) of the pulse-labeled Arg-DHFR<sup>ts</sup> (but not of Met-DHFR<sup>ts</sup>) was degraded during the 10-min pulse at 37 °C, as could be seen from comparing the UPR-normalized amounts of Arg-DHFR<sup>ts</sup> at time 0 (the beginning of the chase) at 23 °C and 37 °C (Fig. 4A). In contrast with a conventional pulse-chase assay, the use of UPR in a setting where a protein can be made either short-lived or long-lived allows the detection and measurement of proteolysis not only during the chase but during the pulse as well [23]. The extent of degradation of a protein during the pulse is denoted as ID<sup>v</sup> [23]. In the present context, the variable ID<sup>v</sup> is defined as the extent of degradation of a radiolabeled short-lived protein (Arg-DHFR<sup>ts</sup> at 37 °C) at the end of a pulse relative to the amount of a nearly identical protein (Met-DHFR<sup>ts</sup> at 37 °C) that is long-lived under the same conditions.

A large fraction, but not all, of the Arg-DHFR<sup>ts</sup> degradation in mouse cells was carried out by the N-end rule pathway. This could be seen by comparing the UPR-based decay curves of Met-DHFR<sup>ts</sup> and Arg-DHFR<sup>ts</sup> at 37 °C (Fig. 3B lanes a–d vs. lanes e–h and Fig. 4A). Specifically, not only Arg-DHFR<sup>ts</sup> but also Met-DHFR<sup>ts</sup> were metabolically unstable at 37 °C in mouse cells, the corresponding  $t_{0.5}^{0-60}$  being ≈ 10 min and ≈ 48 min, respectively (Fig. 4A). As Met is a stabilizing residue in the

N-end rule [23,30], we infer that Arg-DHFR<sup>ts</sup> was targeted not only by the N-end rule pathway, but also by another proteolytic system, which recognized a structural feature that resulted from a conformational perturbation of the DHFR<sup>ts</sup> moiety at 37 °C. This aspect of Arg-DHFR<sup>ts</sup> acted as a degron in mouse cells but not in yeast, inasmuch as the same protein was targeted exclusively by the N-end rule pathway in *S. cerevisiae* at 37 °C (Fig. 2A lanes g–i vs. lanes j–l). The nature of this degradation signal in the conformationally perturbed (at 37 °C) DHFR<sup>ts</sup> moiety is not known.

#### Inhibition of both degradation signals of Arg-DHFR<sup>ts</sup>-ha by MTX in mouse cells

Mouse L cells expressing Met-DHFR<sup>ts</sup> (dha-Ub-Met-DHFR<sup>ts</sup>) or Arg-DHFR<sup>ts</sup> (dha-Ub-Arg-DHFR<sup>ts</sup>) were incubated at 37 °C for 30 min with 20 μM MTX in the medium, followed by a 10-min pulse with [<sup>35</sup>S]methionine/cysteine, a chase, and SDS/PAGE analysis. The degradation of Met-DHFR<sup>ts</sup> at 37 °C, which is mediated by a non-N-degron in the conformationally perturbed DHFR<sup>ts</sup> moiety (see above), was virtually completely inhibited by MTX (Fig. 3B lanes a–d vs. Fig. 3C lanes a–d and Fig. 4B). Specifically, the  $t_{0.5}^{0-60}$  of Met-DHFR<sup>ts</sup> at 37 °C was ≈ 48 min in the absence of MTX and > 10 h in the presence of MTX (Fig. 4B). The much more rapid degradation of Arg-DHFR<sup>ts</sup> carried out by both the N-end rule pathway and the other proteolytic pathway was strongly but incompletely inhibited by MTX (Fig. 3B lanes e–h vs. Fig. 3C lanes e–h and Fig. 4B). Specifically, in the absence of MTX, the  $t_{0.5}^{0-60}$  of Arg-DHFR<sup>ts</sup> was ≈ 10 min (ID<sup>10</sup> of 36%) (Fig. 4B). In the presence of MTX, these variables changed to  $t_{0.5}^{0-60} \approx 37$  min and ID<sup>10</sup> of 29% (Fig. 4B).

A comparison of the corresponding decay curves (Fig. 4), made more reliable by the increased accuracy of a UPR-based pulse-chase assay, suggested one reason for the leaky inhibition of the N-degron of Arg-DHFR<sup>ts</sup> by MTX. Specifically, the ID<sup>10</sup> of Met-DHFR<sup>ts</sup> (i.e. the extent of Met-DHFR<sup>ts</sup> degradation during the pulse) was < 10% in the absence of MTX, whereas

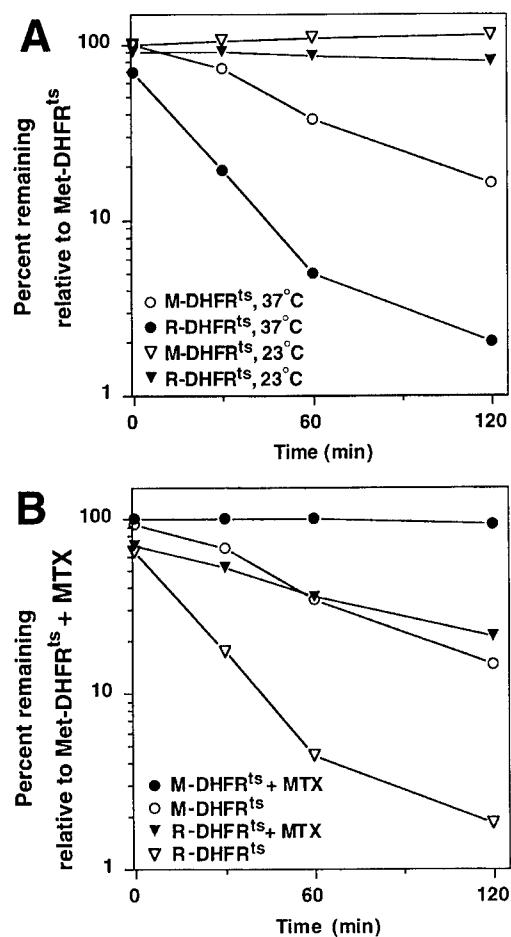


Fig. 4. Decay curves of Met-DHFR<sup>ts</sup> and Arg-DHFR<sup>ts</sup> at 23 °C and 37 °C in mouse cells, in the absence (A) and presence of MTX (B). These curves were derived from the UPR-based electrophoretic data (Figs 3 and 5 and analogous evidence), as described in Materials and methods.

the ID<sup>10</sup> of Arg-DHFR<sup>ts</sup> under the same conditions was ≈ 36%. In other words, more than one-third of labeled Arg-DHFR<sup>ts</sup> molecules were degraded during the 10-min pulse (Fig. 4B). Moreover, the ID<sup>10</sup> of Arg-DHFR<sup>ts</sup> remained high (≈ 30%) even in the presence of MTX (Fig. 4B). This result suggested that a large fraction of the newly formed Arg-DHFR<sup>ts</sup> molecules was targeted for degradation by the N-end rule pathway before the conformation of Arg-DHFR<sup>ts</sup> was mature enough to allow high-affinity binding of MTX (see the Discussion).

#### An exposed N-terminus renders Arg-DHFR<sup>ts</sup> constitutively short-lived

To address the relative contributions of the first (N-terminal Arg) and the second (internal Lys) determinant of the N-degron in Arg-DHFR<sup>ts</sup>, we extended its N-terminus with a 42-residue, *Escherichia coli* Lac repressor-derived sequence, termed e<sup>ΔK</sup> [extension (e) lacking lysines (ΔK)] [16]. Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup> (Ub-Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup>) (Fig. 1, construct IV), which bore a stabilizing N-terminal residue, was long-lived at 23 °C in mouse cells ( $t_{0-60}^{0-60} > 10$  h) (Fig. 5A, lanes a-c). Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup> (Ub-A<sup>ΔK</sup>-e<sup>ΔK</sup>-DHFR<sup>ts</sup>) (Fig. 1, construct V), which bore a destabilizing N-terminal residue, was short-lived at 23 °C ( $t_{0.5}^{0-60} \approx 45$  min; Fig. 5A, lanes d-f), in contrast with the otherwise identical Arg-DHFR<sup>ts</sup> that lacked the e<sup>ΔK</sup> extension and was long-lived at 23 °C (Fig. 3A, lanes e-h). Note that Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup> was short-lived at 23 °C, in spite of the fact that the e<sup>ΔK</sup> extension did not contribute additional Lys residues to Arg-DHFR<sup>ts</sup>, which was long-lived at 23 °C.

At 37 °C, Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup> was approximately as short-lived as Met-DHFR<sup>ts</sup>; the degradation of these proteins was carried out by a non-N-end rule pathway that targeted a degron (d<sub>N</sub>) in the conformationally perturbed DHFR<sup>ts</sup> moiety (Fig. 5B, lanes a-c; compare with Fig. 3B, lanes a-d). In contrast, the degradation of Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup> at 37 °C was much faster than that of Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup>, the difference in rate being due to the targeting of Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup> by both the N-end rule pathway

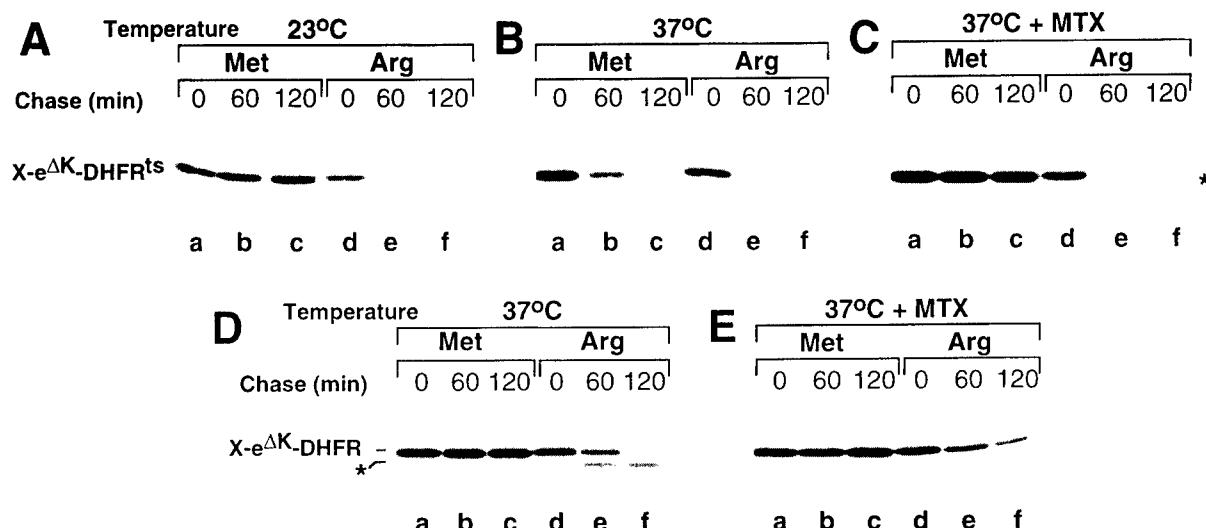


Fig. 5. Effects of temperature and MTX on the degradation of N-terminally extended DHFR<sup>ts</sup> and DHFR in mouse cells. (A) Lanes a-c, mouse L-cells transiently expressing Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup> (Fig. 1, construct IV) were labeled at 23 °C for 10 min with [<sup>35</sup>S]methionine/cysteine, followed by a chase, also at 23 °C, for 60 and 120 min in the presence of cycloheximide, extraction, immunoprecipitation, and SDS/PAGE analysis. Lanes d-f, same as lanes a-c but with Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>. (B) Lanes a-f, same as lanes a-f in (A), but the pulse-chase was carried out at 37 °C. (C) Lanes a-c, same as lanes a-c in (B), but the pulse-chase of Met-e<sup>ΔK</sup>-DHFR<sup>ts</sup> was carried out in the presence of 20 μM MTX in the growth medium. Lanes d-f, same as lanes a-c, but with Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>. (D) Lanes a-c, same as lanes a-c in (B), but with Met-e<sup>ΔK</sup>-DHFR (the wild-type DHFR moiety). Lanes d-f, same as lanes a-c, but with Arg-e<sup>ΔK</sup>-DHFR. The asterisk indicates a cleavage product derived from a minority of short-lived Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup> and Arg-e<sup>ΔK</sup>-DHFR molecules during their targeting by the N-end rule pathway (see the main text). (E) Lanes a-c, same as lane a-c in (D), but in the presence of 20 μM MTX. Lanes d-f, same as lanes a-c but with Arg-e<sup>ΔK</sup>-DHFR.

and the  $d_x$ -recognizing pathway (Fig. 5B, lanes d–f). A minor proteolytic fragment of Arg-e $^{\Delta K}$ -DHFR $^{ts}$ , visible in Fig. 5B–E (lanes d–f), was specific to N-end rule substrates bearing an e $^{\Delta K}$ -type extension; an analogous fragment was also observed with X-e $K$ - $\beta$ gal-based N-end rule substrates [16].

We also examined the metabolic fate of Met-e $^{\Delta K}$ -DHFR (Ub-Met-e $^{\Delta K}$ -DHFR) and Arg-e $^{\Delta K}$ -DHFR (Ub-Arg-e $^{\Delta K}$ -DHFR), which were identical with the substrates above except that they bore the wild-type mouse DHFR moiety (Fig. 1, constructs VI and VII). Met-e $^{\Delta K}$ -DHFR was long-lived at both 23 °C and 37 °C in mouse cells (Fig. 5D, lanes a–c, and data not shown), indicating that the heat-activated  $d_x$ -type degron of substrates containing the DHFR $^{ts}$  moiety (e.g. Fig. 4A) resulted from the Pro66 gr; Leu66 mutation that yielded this moiety. The degradation of Arg-e $^{\Delta K}$ -DHFR was mediated exclusively by the N-end rule pathway (Fig. 5D, lanes d–f vs. lanes a–c), unlike the double-pathway degradation of Arg-e $^{\Delta K}$ -DHFR $^{ts}$  at 37 °C (Fig. 5C, lanes d–f). In contrast with the results with Arg-DHFR $^{ts}$  and Met-DHFR $^{ts}$ , the presence of MTX did not result in a significant stabilization of Arg-e $^{\Delta K}$ -DHFR (Fig. 5D lanes d–f vs. Figure 5E lanes d–f).

## DISCUSSION

We report the following results.

(a) Arg-DHFR $^{ts}$  is long-lived at 23 °C but short-lived at 37 °C in both *S. cerevisiae* and mouse cells.

(b) In yeast, Arg-DHFR $^{ts}$  is degraded (at 37 °C) exclusively by the N-end rule pathway, whereas in mouse cells the same protein at the same temperature is degraded by another proteolytic pathway as well. The corresponding mouse-specific degron is active at 37 °C but inactive at 23 °C. This degradation signal is a currently unknown feature of the perturbed conformation of the DHFR $^{ts}$  moiety at 37 °C, which differs from the wild-type DHFR moiety by a Pro → Leu alteration at position 66.

(c) MTX, a low- $M_r$  ligand of DHFR ( $K_d \approx 10$  pm), inhibits the degradation of Arg-DHFR $^{ts}$  (at 37 °C) nearly completely in yeast and partially in mouse cells.

(d) The ability of MTX to inhibit the degradation of Arg-DHFR $^{ts}$  *in vivo* is retained when Arg-DHFR $^{ts}$  is utilized as a portable degron that confers metabolic instability on linked unrelated proteins in yeast.

(e) Using the UPR technique [23] to determine the metabolic fate of Arg-DHFR $^{ts}$  in mouse cells, we found that  $\approx 36\%$  of the labeled Arg-DHFR $^{ts}$  molecules are degraded during the 10-min pulse, and that the presence of MTX decreases this fraction only slightly, to  $\approx 30\%$ . This finding accounts for the leaky inhibition of Arg-DHFR $^{ts}$  degradation by MTX in mouse cells. Specifically, this finding suggests that a large fraction of the nascent Arg-DHFR $^{ts}$  is targeted, in mouse cells, by the N-end rule pathway before the conformation of the DHFR $^{ts}$  moiety is mature enough [28] to allow high-affinity binding of MTX. That the MTX-mediated inhibition of degradation of Arg-DHFR $^{ts}$  in yeast is much less leaky than in mouse cells is but one difference in the detailed properties of the yeast and mammalian N-end rule pathways [see also item (f)].

(f) When the N-terminus of Arg-DHFR $^{ts}$  was extended with a 42-residue lysine-lacking extension, termed e $^{\Delta K}$ , the N-degron of the resulting Arg-e $^{\Delta K}$ -DHFR $^{ts}$  was active at both 23 °C and 37 °C, unlike the mouse-specific degron, which remained inactive at 23 °C. Moreover, the degradation of Arg-e $^{\Delta K}$ -DHFR $^{ts}$ , in contrast with that of Arg-DHFR $^{ts}$ , could not be inhibited by MTX, suggesting that the inactivity of the N-degron in Arg-DHFR $^{ts}$  at 23 °C results at least in part from inaccessibility

of its N-terminal arginine to the targeting complex of the N-end rule pathway. Arg-e $^{\Delta K}$ -DHFR, which bore the wild-type DHFR moiety, was also short-lived in mouse cells. Previously, the same protein was found to be long-lived in yeast, unless the e $^{\Delta K}$  extension was replaced by an otherwise identical extension, termed e $K$ , which contained Lys residues that functioned as the second determinant of the N-degron [16].

Potential applications of a heat-inducible, MTX-suppressible portable N-degron include its use to produce conditional mutants in homeothermic animals such as mouse. One difficulty in the current approaches to this problem [31–33] is that either a deletion or transcriptional repression of a gene of interest leaves the previously produced gene product unperturbed. If this protein is long-lived and resides in a non-dividing cell, there may be a considerable phenotypic lag between the conditionally introduced genetic change and the actual inactivation or disappearance of a corresponding protein. A degron that can be regulated by a cell-penetrating ligand can be employed to address this problem. Since chronic administration of MTX is toxic to mammals, it will be necessary to construct an analogous DHFR-based N-degron that uses *E. coli* DHFR and is inhibited by trimethoprim, which binds tightly to the *E. coli* but not the mouse DHFR [34]. If a portable N-degron of this kind could be constructed, combining it with a conditional repression of a gene of interest should yield better methods for producing conditional mutants in homeothermic animals.

MTX, through its binding to the substrate pocket of DHFR, stabilizes DHFR conformation [35,36]. This effect was employed in several studies with DHFR as a reporter protein. For example, the binding of MTX to DHFR can preclude, under certain conditions, the translocation of a DHFR-containing fusion protein across biological membranes [37,38]. MTX can partially protect DHFR against *in vitro* proteolysis by thermolysin [38,39]. MTX has also been shown to block the degradation of a targeted multiubiquitylated Arg-e $K$ -DHFR in reticulocyte extract [27]. In the latter study, the N-end rule pathway in reticulocyte extract was presented with a prefolded wild-type DHFR moiety. By contrast, in the present work with intact cells, the same pathway could target a DHFR-based N-end rule substrate immediately after (and possibly even during) its synthesis. Further analysis is required to clarify the apparent discrepancy between the previously observed inhibition of degradation of Arg-e $K$ -DHFR by MTX *in vitro* (in reticulocyte extract) [27] and the absence of a comparable effect of MTX *in vivo* on this class of substrates, in which N-terminal Arg is constitutively exposed. One difference between the two settings is the presentation of prefolded DHFR moieties in reticulocyte extract compared with a kinetic competition between the folding of newly formed DHFR moieties and their targeting for degradation by the N-end rule pathway.

The use of the UPR technique not only increased the overall accuracy of pulse-chase assays with mouse cells but also showed that  $\approx 30\%$  of the newly made Arg-DHFR $^{ts}$  was degraded during the pulse. The finding that the fraction of Arg-DHFR $^{ts}$  degraded during the pulse decreased only slightly in the presence of MTX strongly suggests that most of the labeled Arg-DHFR $^{ts}$  is conformationally immature shortly after synthesis, and therefore cannot form a high-affinity complex with MTX. The Pro → Leu alteration at position 66 that yielded DHFR $^{ts}$  occurred in the region of DHFR that interacts with the aromatic ring of MTX [35], and is likely to have resulted in a decreased affinity of DHFR for MTX. Nonetheless, the test proteins containing the DHFR $^{ts}$  moiety were efficiently retained on an MTX affinity column (data not shown), and their *in vivo* degradation could be specifically inhibited by MTX.

A model of targeting by the N-end rule pathway that accounts for our findings with the heat-inducible N-degron assumes that the main cause of metabolic stability of Arg-DHFR<sup>ts</sup> at 23 °C is not the absence of a sterically accessible Lys residue within the folded DHFR globule, but rather a steric inaccessibility of the N-terminal Arg residue. In this model, the activation of the previously cryptic N-degron of Arg-DHFR<sup>ts</sup> at 37 °C is caused by an increased exposure and flexibility of a region bearing the N-terminal Arg, a transition that leads to the binding of Arg by the targeting complex of the N-end rule pathway. According to this interpretation, MTX stabilizes Arg-DHFR<sup>ts</sup> against degradation at 37 °C not by precluding the conformational mobilization of internal Lys residues, but by precluding a temperature-mediated increase in the exposure of the N-terminal Arg. In addition to being consistent with the available evidence, this model accounts for the otherwise puzzling result that Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>-ha (which differs from Arg-DHFR<sup>ts</sup> by the presence of e<sup>ΔK</sup>) is short-lived at both 23 °C and 37 °C, whereas Arg-DHFR<sup>ts</sup> is short-lived only at 37 °C. Indeed, the e<sup>ΔK</sup> extension would result in a temperature-independent enhanced exposure of N-terminal Arg, a change that would be, in this model, sufficient for the temperature-independent activity of the N-degron in Arg-e<sup>ΔK</sup>-DHFR<sup>ts</sup>. The earlier model proposed by Dohmen *et al.* [19], which presumed that the inactivity of the N-degron in Arg-DHFR<sup>ts</sup> at 23 °C was caused by the absence of sterically accessible Lys residues, cannot account for the above result.

How can we verify the Arg-exposure model? One way is suggested by the existence, in both yeast and mammals, of the enzymes N-terminal amidase (Nt-amidase) and Arg-tRNA-protein transferase (R-transferase), which chemically modify specific (sterically exposed) N-terminal residues of proteins in the cytosol. In *S. cerevisiae*, the *NTA1*-encoded Nt-amidase deamidates N-terminal Asn or Gln; the *ATE1*-encoded R-transferase conjugates Arg to N-terminal Asp or Glu (reviewed in [10]). A test of the model would involve the expression of, for example, Asn-DHFR<sup>ts</sup> (identical to Arg-DHFR<sup>ts</sup> except for the presence of N-terminal Asn) in yeast cells that lack Ubr1p, a component of the N-end rule pathway that recognizes primary destabilizing N-terminal residues such as Arg. (The absence of Ubr1p would preclude the degradation of test proteins.) The model predicts that at 23 °C the N-terminal Asn of Asn-DHFR<sup>ts</sup> would be deamidated inefficiently or not at all, in contrast with what happens at 37 °C. This can be verified by isolating Asn-DHFR<sup>ts</sup> from cells incubated at either 23 °C or 37 °C, and determining its N-terminal residue. If the model is correct, the N-terminal sequence must begin largely or entirely with Asn at 23 °C, but would become Arg-Asp... at 37 °C. (In the latter case, the N-terminal Asp, produced from Asn, would be arginylated by R-transferase.) Experimental verification of this model is under way.

## ACKNOWLEDGEMENTS

We thank Michel Ghislain, Ailsa Webster and Martin Gonzalez for helpful discussions and advice, and Lawrence Peck for comments on the manuscript. This work was supported by the grants to A.V. from the National Institutes of Health (DK39520 and GM31530). F.L. and J.A.J. were supported by fellowships from the Swiss National Fund for Research and the American Cancer Society, respectively.

## REFERENCES

- Haas, A.J. & Siepmann, T.J. (1997). Pathways of ubiquitin conjugation. *FASEB J.* **11**, 1257–1268.
- Hershko, A. (1997). Roles of ubiquitin-mediated proteolysis in cell cycle control. *Curr. Op. Cell. Biol.* **9**, 788–799.
- King, R.W., Deshaies, R.J., Peters, J.M. & Kirschner, M.W. (1996). How proteolysis drives the cell cycle. *Science* **274**, 1652–1659.
- Varshavsky, A. (1997). The ubiquitin system. *Trends Biochem. Sci.* **22**, 383–387.
- Wilkinson, K.D. (1997). Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *FASEB J.* **11**, 1245–1256.
- Pickart, C.M. (1997). Targeting of substrates to the 26S proteasome. *FASEB J.* **11**, 1055–1066.
- Murray, A. & Hunt, T. (1993) *The Cell Cycle*. Freeman, New York.
- Varshavsky, A. (1991). Naming a targeting signal. *Cell* **64**, 13–15.
- Bachmair, A., Finley, D. & Varshavsky, A. (1986). *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186.
- Varshavsky, A. (1997). The N-end rule pathway of protein degradation. *Genes to Cells* **2**, 13–28.
- Hicke, L. (1997). Ubiquitin-dependent internalization and down-regulation of plasma membrane proteins. *FASEB J.* **11**, 1215–1226.
- Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439.
- Scheffner, M., Smith, S. & Jentsch, S. (1998) In *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J.R. & Finley, D., eds), pp. 65–98. Plenum Press, New York.
- Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380.
- Hilt, W. & Wolf, D.H. (1996). Proteasomes: destruction as a programme. *Trends. Biochem. Sci.* **21**, 96–102.
- Bachmair, A. & Varshavsky, A. (1989). The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032.
- Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K. & Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576–1583.
- Hill, C.P., Johnston, N.L. & Cohen, R.E. (1993). Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl. Acad. Sci. USA* **90**, 4136–4140.
- Dohmen, R.J., Wu, P. & Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276.
- Sikorski, R.S. & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. *Genetics* **122**, 19–27.
- Johnson, E.S., Bartel, B.W. & Varshavsky, A. (1992). Ubiquitin as a degradation signal. *EMBO J.* **11**, 497–505.
- Johnson, E.S., Ma, P.C.M., Ota, I.M. & Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270**, 17442–17456.
- Lévy, F., Johnsson, N., Rümenapf, T. & Varshavsky, A. (1996). Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907–4912.
- Ghislain, M., Dohmen, R.J., Lévy, F. & Varshavsky, A. (1996). Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 4884–4899.
- Baker, R.T. & Varshavsky, A. (1991). Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **87**, 2374–2378.
- Mosteller, R.D. & Goldstein, B.E. (1984). A mathematical model that applies to protein degradation and posttranslational processing of proteins and to analogous processes for other molecules in non-growing and exponentially growing cells. *J. Theor. Biol.* **108**, 597–621.
- Johnston, J.A., Johnson, E.S., Waller, P.R.H. & Varshavsky, A. (1995). Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. *J. Biol. Chem.* **270**, 8172–8178.
- Frieden, C. (1990). Refolding of *Escherichia coli* dihydrofolate reductase: sequential formation of substrate-binding sites. *Proc. Natl. Acad. Sci. USA* **87**, 4413–4416.
- Spence, J., Sadis, S., Haas, A.L. & Finley, D. (1995). A ubiquitin mutant

- with specific defects in DNA repair and multiubiquitination. *Mol. Cell. Biol.* **15**, 1265–1273.
30. Gonda, D.K., Bachmair, A., Wünning, I., Tobias, J.W., Lane, W.S. & Varshavsky, A. (1989). Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712.
  31. Gu, H., Marth, J.D., Orban, P.C., Moosmann, H. & Rajewski, K. (1994). Deletion of a DNA polymerase  $\beta$  gene segment in T cells using cell type-specific gene targeting. *Science* **265**, 103–106.
  32. No, D., Yao, T.-P. & Evans, R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **93**, 3346–3351.
  33. St-Onge, L., Furth, P.A. & Gruss, P. (1996). Temporal control of the Cre recombinase in transgenic mice by a tetracycline-responsive promoter. *Nucl. Acids Res.* **24**, 3975–3877.
  34. Matthews, D.A., Bolin, J.T., Burridge, J.M., Filman, D.J., Volz, K.W. & Kraut, J. (1985). Dihydrofolate reductase: the stereochemistry of inhibitor selectivity. *J. Biol. Chem.* **260**, 392–399.
  35. Oefner, C., D'Arcy, A. & Winkler, F.K. (1988). Crystal structure of human dihydrofolate reductase complexed with folate. *Eur. J. Biochem.* **174**, 377–385.
  36. Thillet, J., Absil, J., Stone, S.R. & Pictet, R. (1988). Site-directed mutagenesis of mouse dihydrofolate reductase: mutants with increased resistance to methotrexate and trimethoprim. *J. Biol. Chem.* **263**, 12500–12508.
  37. Arkowitz, R.A., Joly, J.C. & Wickner, W. (1992). Translocation can drive the unfolding of a preprotein domain. *EMBO J.* **12**, 243–253.
  38. Eilers, M. & Schatz, G. (1986). Binding of a specific ligand inhibits import of a purified precursor into mitochondria. *Nature* **322**, 228–232.
  39. Klingenberg, O. & Olsnes, S. (1996). Ability of methotrexate to inhibit translocation to the cytosol of dihydrofolate reductase fused to diphtheria toxin. *Biochem. J.* **313**, 647–653.
  40. Madura, K. & Varshavsky, A. (1994). Degradation of G $\alpha$  by the N-end rule pathway. *Science* **265**, 1454–1458.
  41. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A. & Seidman, J.G. & Struhl, K. (1996) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York.
  42. Madura, K., Dohmen, R.J. & Varshavsky, A. (1993). N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.* **268**, 12046–12054.

# Codominant interference, antieffectors, and multitarget drugs

(pharmacology/cancer/codominance/selectivity)

ALEXANDER VARSHAVSKY†

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Alexander Varshavsky, December 19, 1997

**ABSTRACT** The insufficient selectivity of drugs is a bane of present-day therapies. This problem is significant for antibacterial drugs, difficult for antivirals, and utterly unsolved for anticancer drugs, which remain ineffective against major cancers, and in addition cause severe side effects. The problem may be solved if a therapeutic agent could have a multitarget, combinatorial selectivity, killing, or otherwise modifying, a cell if and only if it contains a predetermined set of molecular targets and lacks another predetermined set of targets. An earlier design of multitarget drugs [Varshavsky, A. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3663–3667] was confined to macromolecular reagents such as proteins, with the attendant difficulties of intracellular delivery and immunogenicity. I now propose a solution to the problem of drug selectivity that is applicable to small ( $\leq 1$  kDa) drugs. Two ideas, codominant interference and antieffectors, should allow a therapeutic regimen to possess combinatorial selectivity, in which the number of positively and negatively sensed macromolecular targets can be two, three, or more. The nature of the effector and interference moieties in a multitarget drug determines its use: selective killing of cancer cells or, for example, the inhibition of a neurotransmitter-inactivating enzyme in a specific subset of the enzyme-containing cells. The *in vivo* effects of such drugs would be analogous to the outcomes of the Boolean operations “and,” “or,” and combinations thereof. I discuss the logic and applications of the antieffector and interference/codominance concepts, and the attendant problem of pharmacokinetics.

The many successes of pharmacology (1, 2) do not include the problem of cancer. Major human cancers are incurable once they have metastasized. A few relatively rare cancers, such as testicular carcinoma in men, Wilms' kidney tumor, and some leukemias in children, can often be cured through chemotherapy but require cytotoxic treatments of a kind that cause severe side effects and are themselves carcinogenic (2).

The main reason for the failure of cytotoxic therapies is their insufficient selectivity for tumors. For example, treatments with radiation or alkylating agents perturb many functions that are common to all cells. The more selective cytotoxic drugs, for instance, methotrexate, taxol, and etoposide, perturb the functions of specific macromolecular targets (dihydrofolate reductase, microtubules, and topoisomerase II), but these targets are present in both normal and malignant cells (1, 2). Hence the low therapeutic index of anticancer drugs and their systemic toxicity at clinically relevant doses. Because the mitotic activity of cells in a tumor is often lower than the mitotic activity of normal cells in self-renewing tissues such as the bone marrow (3), one might not have expected these drugs to work at all—to have any preference for the killing of cancer cells. That such preference actually exists stems in part from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/952094-6\$2.00/0  
PNAS is available online at <http://www.pnas.org>.

the fact that tumor cells are often perturbed by their mutations into stress-hypersensitive states. Consequently, these tumor cells die an apoptotic death at the level of a drug-imposed metabolic stress that induces apoptosis in some but not in most of the organism's normal cells (3).

With some cancers, cytotoxic therapies are ineffective from the beginning. In other cases, these therapies yield a partial, sometimes clinically complete, but almost invariably transient remission of a cancer, in part because these treatments select for tumor cell variants that retain tumorigenicity but are more resistant to either apoptosis *per se* or a drug that induces apoptosis. Because a significant increase in the drug or radiation dosage is precluded by their low therapeutic index, these therapies become ineffective when resistant clones of malignant cells, selected by a drug treatment, present themselves as a cancer recurrence.

The failure of small cytotoxic drugs to produce a cure for cancer has given rise to other strategies, in particular the insightful suggestion that solid tumors can be targeted by selectively inhibiting neovascularization, a process that these tumors depend on for growing to a clinically significant size (4). Another approach, immunotoxins, involves the linking of a toxin to a ligand such as an antibody or a growth factor that binds to a target on the surface of tumor cells (5). Among the limitations of present-day immunotoxins is their incapacity, on entering a cell, to adjust their toxicity in response to the intracellular protein composition. Yet another approach is to enhance the ability of the immune system to identify and selectively destroy tumor cells. The current revival of this strategy holds the promise of a rational and curative treatment (6). Given the complicated regimens and unsolved problems of immunotherapies, it is clear that this and other recent approaches (7, 8) are motivated in part by the perception that small-drug pharmacology, so successful against bacterial infections, is unlikely to prove effective against cancer. In contrast to this view, the premise of the strategy described in the present work is that small anticancer drugs may become curative and free of severe side effects if a way is found to confer on these compounds a multitarget, combinatorial selectivity.

Most cancers are monoclonal: cell lineages of both the primary tumor and the metastases originate from a single founder cell. This cell is a breakthrough descendant of a cell lineage that has been accumulating mutations for some time, often in proximity to other neoplastic but still nonmetastatic cell lineages within an indolent proliferative lesion such as a benign tumor (9, 10). Given the monoclonality of a cancer, cells of both the primary tumor and the metastases share the initial mutations that yielded the founder cell, even if these cells differ at other loci that accumulated mutations in the course of the later tumor progression. Some of the early mutations are in genes that encode tumor suppressors (9, 11, 12). In most cancers, both alleles of a tumor suppressor gene are inactivated, sometimes through deletions that encompass the gene on the two homologous chromosomes. Thus, a monoclonal cancer, although heterogeneous genetically,

Abbreviation: IC, interference/codominance.

†To whom reprint requests should be addressed at: Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. e-mail: [avarsh@cco.caltech.edu](mailto:avarsh@cco.caltech.edu).

always contains a set of founder mutations that is shared by all of its cells.

A drug that kills a cell only if it lacks a specific macromolecular target would distinguish tumor cells from many other cells of an organism, provided that the target is a product of a gene that had been deleted or inactivated in this cancer at the stage of its founder cell. Such a drug may be especially selective against cancers that lack a gene for a ubiquitously expressed tumor suppressor, for example, the retinoblastoma (Rb) protein (11, 12). An example of the negative-target approach is the use of a mutant adenovirus that replicates selectively in human cancer cells lacking the tumor suppressor p53 and has been shown to kill these cells in a model setting (8).

However, other tumor suppressors may not be expressed at comparable levels in most cells. A drug that kills a cell if it lacks a nonubiquitous tumor suppressor would be toxic to a subset of normal cells as well. This problem could be reduced through the use of a drug that is toxic only if a cell lacks two specific macromolecules, termed negative targets. Two judiciously chosen negative targets may, together, suffice to distinguish all of the cancer cells from all of the organism's normal cells. If they do not, a third negative target that had been deleted or rendered defective in a given cancer can be employed as well. This strategy requires a drug that possesses the ability to kill a cell if it lacks two or more of the predetermined targets, but would spare a cell containing either one of these targets.

Other changes in a founder cell may involve a missense mutation, an amplification and overexpression, an ectopic expression, or a translocation/fusion of a specific protooncogene such as, for example, *Ras* or *Myc* (9, 10, 13). A single oncprotein may not be a unique enough target by itself, for reasons similar to those described above in the context of negative targets. However, a combination of two or more distinct oncproteins that were either mutated or inappropriately expressed in the founder cell can be employed to formulate the unique multiprotein signature of a specific cancer that comprises both positive and negative targets.

These considerations suggest that a conditionally cytotoxic therapeutic regimen that is exquisitely specific for a given cancer, and therefore would eliminate it without significant side effects, must possess, in most cases, a multitarget, combinatorial (positive/negative) selectivity of the kind defined above. Conversely, even an informed choice of the molecular target for a single-target drug may not suffice to define unambiguously the cell type to be eliminated. Note that simply combining two single-target drugs against two different targets in a multidrug regimen would not yield a multitarget selectivity, because the two drugs together would perturb not only cells containing both targets but also cells containing either one of the targets.

Although the problem of insufficient selectivity is not as acute with noncytotoxic drugs, it is relevant to them as well. Among the multitude of examples are side effects of therapies with antipsychotic agents. The side effects are caused in part by the insufficient molecular specificity of drugs, which is exemplified by the ability of antidepressants that inhibit monoamine oxidase to perturb other proteins as well (2). This difficulty will continue to abate with the development of more specific single-target inhibitors. But an entirely distinct, major, and unsolved problem with inhibitors as drugs is the current impossibility of restricting their action to a specific subset of cells among those that contain the inhibitor's target. For example, even an exquisitely specific inhibitor of a clinically relevant enzyme is likely to have side effects, because the target enzyme is present, in most cases, not only in the cells where its inhibition is clinically beneficial but also in the cells where its inhibition is physiologically inappropriate. The present work describes a possible solution of this problem.

A previously proposed approach to designing multitarget drugs utilized degradation signals (degrons) and analogous signals that

exhibit the property of codominance (14, 15). As a result, this strategy was confined to macromolecular reagents such as proteins, with the attendant problems of immunogenicity, extravasation, and intracellular delivery. The latter difficulty is especially significant, because either gene-therapy or direct-delivery methods for introducing large molecules into cells work reasonably well with cells in culture but are still inefficient with cells in an intact organism. The challenge, then, is to attain a multitarget, combinatorial selectivity in the setting of small ( $\leq 1$  kDa) drugs, where the immunogenicity and delivery problems are less severe. A solution, described below, invokes a modification of the earlier idea of codominant interference (14) in conjunction with the new concept of antieffectors. This solution is applicable to either cytotoxic or noncytotoxic therapies.

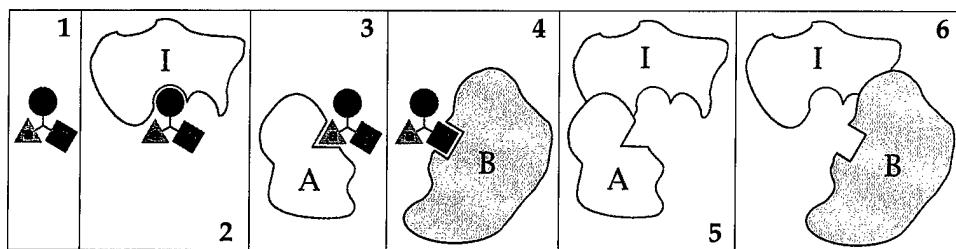
## RESULTS AND DISCUSSION

**Multitarget Compounds Specific for Negative Targets: The Concept of Codominant Interference.** Previous work (14) suggested that the property of codominance, characteristic of degradation signals (degrons) and many other signals in biopolymers, can be employed to design protein-based reagents that possess multitarget, combinatorial selectivity of the kind defined above. Codominance refers to the ability of two or more signals in the same molecule to function independently and not to interfere with each other. It is shown below that a distinct version of the interference/codominance (IC) concept (14) is applicable to small ( $\leq 1$  kDa) compounds. Consider a reagent containing three small moieties **a**, **b**, and **i**, which can bind, respectively, to three macromolecular targets **A**, **B**, and **I**. Because the moieties **a**, **b**, and **i** are much smaller than the macromolecules **A**, **B**, and **I**, it should be possible to arrange these moieties in the compound **abi** in such a way that the binding of **A** or **B** to **a** or **b** would preclude, through steric hindrance, the binding of moiety **i** to **I** (Fig. 1).

That the interactions of a small bipartite compound with its two macromolecular ligands can be made mutually exclusive is expected from basic physicochemical considerations. This has also been demonstrated directly, in a context unrelated to the present discussion. When lisinopril, an inhibitor of the angiotensin-converting enzyme (ACE), was connected, via an 11-atom linker, to the biotin moiety, the resulting bivalent compound could bind to and inhibit ACE in the absence but not in the presence of the biotin-binding protein streptavidin (16). Small compounds comprising two linker-connected moieties such as cyclosporin and FK506, which are specific for two macromolecular targets, have previously been employed as *in vivo* dimerization devices, making it possible to bring together two otherwise noninteracting proteins (17). However, the linker moiety of these bipartite compounds was chosen to allow simultaneous interactions with the targets, in contrast to the mutual exclusivity of interactions in the IC approach (Fig. 1).

If the moiety **i** is an inhibitor of an essential cellular enzyme **I**, the presence of the macromolecular targets **A** or **B** in a cell would reduce the inhibition of enzyme **I** by **abi**, because the complexes **abi-A** and **abi-B** would be mutually exclusive with the complex **abi-I** (Fig. 1). Note that **A** and **B** are codominant in their ability to reduce the inhibition of **I** by **abi**. Therefore, there is, formally, no limit on the number of **a**, **b**-like competition modules that can be used to construct an **abi**-like compound whose activity is sensitive to the presence of several distinct macromolecules, called negative targets. The fractional occupancy of the macromolecular targets **A**, **B**, and **I** by the **a**, **b**, and **i** moieties of **abi** would be determined in part by the targets' intracellular concentrations. There are also specific pharmacokinetic constraints on the selectivity of **abi**, an issue discussed below.

A tabulation of the relative toxicities of **abi** for cells that either lack or contain the negative targets **A** and **B** is shown in Fig. 2. It can be seen that **abi** would be relatively nontoxic to three of the four cell types and toxic exclusively to the cells that lack both **A** and **B** (Fig. 2). Thus, the IC concept allows the construction of



the interaction **b-B**. Specifically, the macromolecule **I** in its complex with the small moiety **i** would sterically clash with the macromolecules **A** or **B** if either **A** or **B** is positioned to bind **a** or **b** of **abi** (5 and 6). In the diagram, the interactions **a-A** and **b-B** are also mutually exclusive, but this constraint is not essential. (Note that if the interactions **a-A** and **b-B** were mutually nonexclusive, the compound **abi** would promote the binding of **A** to **B**.) The codominance aspect of the IC concept allows this design to accommodate more than two of the **a**, **b**-like competition modules (not shown). In Figs. 2–4, the **i** moiety is an inhibitor of an essential enzyme **I**. In fact, the only constraint on the identities of **i** and **I** is the requirement for an **i-I** interaction to alter the functional activity of a macromolecule **I**. In other words, the choice of **I** is determined by the intended effect of the (unsequestered) compound **abi** (see the main text).

small compounds that exhibit multitarget selectivity for negative targets. One more idea is required to accomplish the same for positive targets and to link the two strategies.

**Multitarget Compounds Specific for Positive Targets: The Concept of Antieffectors.** Consider a small compound **i\*** that binds to enzyme **I** in the vicinity of its active site, but does not perturb the catalytic activity of **I** toward its physiological substrates (Fig. 3B). Suppose further that the compound **i\***, termed an antiinhibitor, was designed to interfere, sterically, with the binding of an inhibitor **i** to the enzyme's active site while at the same time allowing the binding of physiological substrates. One way to achieve this would be to endow either **i**, or **i\***, or both of them with a set of chemical groups, termed a "bump," whose function is to produce steric hindrance that makes the interactions **i-I** and **i\*-I** mutually exclusive (Fig. 3A and B). A moiety that functions as a bump may also be designed to enhance specific binding of either the inhibitor **i** or the antiinhibitor **i\*** to enzyme **I**, but this consideration is secondary to the bump's essential purpose.

In one application of the antiinhibitor **i\***, it is linked to **c**, a small moiety that can bind to a macromolecular target **C**. The mutual arrangement of **i\*** and **c** in **ci\*** is such that the interactions of **ci\*** with **I** and **C** are mutually exclusive. In the absence of **C**, **ci\*** would compete with **abi** for the binding to

enzyme **I**, thereby partially protecting **I** from inhibition by **abi** (Fig. 3C). This protective effect of **ci\*** would be suppressed in the presence of its macromolecular target **C** (Fig. 3D). In the logic of codominance, discussed above in the context of negative targets, a compound bearing an antiinhibitor moiety **i\*** could contain more than one **c**-like moiety. For example, a compound **cdi\***, whose moieties **c** and **d** can bind, respectively, to the macromolecules **C** and **D**, would reduce the inhibition of enzyme **I** by **abi** only in the absence of both **C** and **D**. Yet another pattern of multitarget selectivity can be produced, in this context, by separating the competition moieties **c** and **d**. The resulting **ci\*** and **di\***, if administered together with **abi**, would reduce the inhibition of enzyme **I** by **abi** if just one of the targets, **C** or **D**, is absent. As shown below, the key merit of the antiinhibitor idea is that it allows the effect of a single inhibitor **i** to be modulated by both negative and positive macromolecular targets.

**On the Difference Between Antieffectors and Antagonists.** The distinctions between substrates, inhibitors, and antiinhibitors were described above. The concept of antieffectors is also relevant to ligand-binding biopolymers other than enzymes. For example, an agonist binds to its receptor and evokes a physiological response. An antagonist binds to a site of the receptor that overlaps with the agonist-binding site, does not

Cell type	Major complexes	Minor complexes	Toxicity of
$A^+ B^+$			Low
$A^+ B^-$			Low
$A^- B^+$			Low
$A^- B^-$			High

FIG. 2. Multitarget selectivity of a compound that utilizes interference/codominance. This diagram tabulates the relative toxicities of the compound **abi** for cells that either lack or contain macromolecular targets **A** and **B**. The **i** moiety of the compound **abi** (see the legend to Fig. 1) inhibits an essential enzyme **I**. The interaction **i-I** is mutually exclusive with the interaction **a-A** and the interaction **b-B**, the macromolecules **A** and **B** being negative targets of **abi**. It is assumed that concentrations of the targets **A** and **B** in cells that contain at least one of them significantly exceed the concentration of **I** (see the main text). In  $A^+ B^+$ ,  $A^+ B^-$ , and  $A^- B^+$  cells, the enzyme **I** would be at most partially inhibited by the **i** moiety of **abi**, because of the competing interactions of **abi** with **A** and/or **B**. By contrast, in  $A^- B^-$  cells, the bulk of **abi** molecules would be available for interaction with **I**, resulting in the selective toxicity of **abi** to these cells. The selectivity pattern of **abi** requires that certain pharmacokinetic conditions are met as well (see the main text). Note that the physiological effects and the uses of **abi**-type compounds are not confined to cytotoxic regimens.

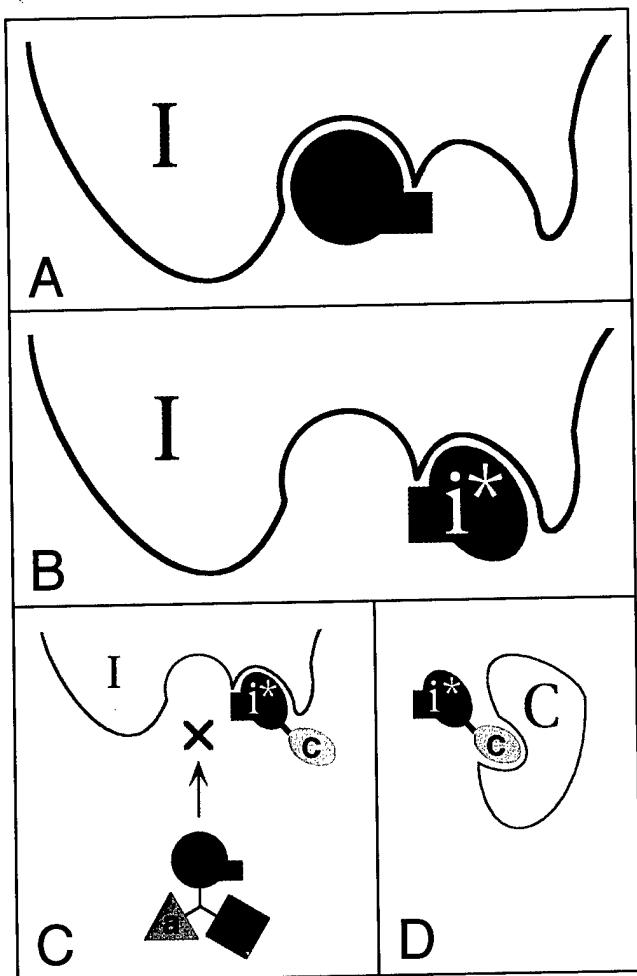


FIG. 3. The antieffector concept. The particular case illustrated here is that of an antiinhibitor  $i^*$ , defined as a compound whose binding to an enzyme  $I$  does not inhibit the activity of  $I$  but does preclude the inhibition of  $I$  by an inhibitor  $i$ . In this example, the antiinhibitor  $i^*$  has the following properties. First, it binds to  $I$  in the vicinity of the  $I$ 's active site, but does not perturb the catalytic activity of  $I$  toward its physiological substrates. Second,  $i^*$ , in its bound state, sterically interferes with the interaction between  $I$  and its inhibitor  $i$ . To implement the second condition, either  $i$ , or  $i^*$ , or both of them bear additional moiety, a "bump," denoted by the rectangular protrusions in  $i$  and  $i^*$ . The function of the bump is to produce steric hindrance that makes the interactions  $i$ - $I$  and  $i^*$ - $I$  mutually exclusive. The inhibitor  $i$  described here and in the main text is a competitive inhibitor, but  $i$  could be a noncompetitive inhibitor as well. An allosteric antiinhibitor, which functions through binding to a remote site of enzyme  $I$ , is yet another possibility. (A) A complex of the enzyme  $I$  with its inhibitor  $i$ . (B) A complex of  $I$  with its antiinhibitor  $i^*$ . Note that the bumps of the bound  $i$  and  $i^*$  spatially overlap. (C) The antiinhibitor  $i^*$  is linked to  $c$ , a small moiety that can bind to a macromolecular target  $C$ . The design of  $ci^*$  is analogous to  $abi$  (Figs. 1 and 2), in that the interactions of  $ci^*$  with  $I$  and  $C$  are mutually exclusive. When  $ci^*$  is bound to  $I$ , the inhibitor  $i$ , shown here as a part of the compound  $abi$  (Figs. 1 and 2), is unable to bind to and inhibit the enzyme  $I$ . (D) A complex between  $ci^*$  and its macromolecular target  $C$ . This complex, being mutually exclusive with the  $ci^*$ - $I$  complex, reduces the ability of  $ci^*$  to protect the enzyme  $I$  from inhibition by  $abi$ .

activate the receptor, and in addition precludes the binding of agonist (1, 2). By contrast, an antieffector, which would be called, in this setting, an antiantagonist, binds to the receptor in such a way that the receptor can still bind, and respond to, the agonist, but cannot bind the antagonist. To this end, either an antagonist, or an antiantagonist, or both must possess a bump, an additional moiety described above in the context of

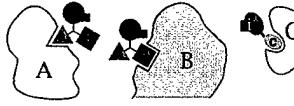
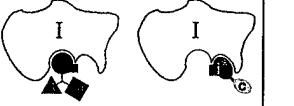
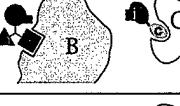
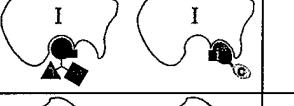
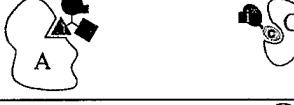
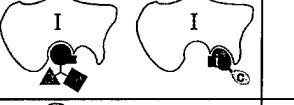
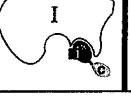
enzymes and antiinhibitors (Fig. 3). The idea of antieffectors is thus distinct from that of antagonists or inhibitors and is new, to the best of my knowledge. I am also not aware of a naturally occurring pair of compounds that satisfy the definition of effectors/antieffectors in a physiologically relevant setting.

**Interference/Codominance and Antieffector in a Regimen That Possesses Combinatorial Selectivity.** Applying the IC and antieffector concepts together yields regimens that possess true combinatorial selectivity, i.e., sensitivity to both negative and positive targets. Consider a population of cells that either contain or lack the macromolecular targets  $A$ ,  $B$ , and  $C$ . Our aim is to devise a treatment that would be toxic to cells that lack  $A$  and  $B$  but contain  $C$  ( $A^- B^- C^+$  cells) and relatively nontoxic to the other cell types (Fig. 4). A regimen of two compounds,  $abi$  (Fig. 3C) and  $ci^*$  (Fig. 3D), has the requisite selectivity, as shown in Fig. 4, which tabulates the outcomes of this treatment for different cell types. Specifically, in the absence of  $C$  (four cell types out of eight), the antiinhibitor-containing  $ci^*$  would compete with the inhibitor-containing  $abi$  for binding to the essential enzyme  $I$ , thereby reducing the inhibition of  $I$  by  $abi$ , and hence reducing the toxicity of  $abi$ . In three other cell types, whose common property is the presence of  $C$  and at least one of the other two targets,  $A$  or  $B$ , the antiinhibitor-containing  $ci^*$  would be largely sequestered by  $C$ , and hence inactive, but the inhibitor-containing  $abi$  would be sequestered as well, by either  $A$  or  $B$ . In only one type of cells, those that lack  $A$  and  $B$  but contain  $C$  ( $A^- B^- C^+$  cells), is the inhibitor  $abi$  fully available for interaction with  $I$ , resulting in higher toxicity (Fig. 4). The differences in the toxicity of  $abi$  to different cell types would be determined by the relative stoichiometries and absolute concentrations of the cellular targets involved ( $A$ ,  $B$ ,  $C$ , and  $I$ ), by the affinities of the moieties  $a$ ,  $b$ ,  $c$ ,  $i$ , and  $i^*$  for these targets, and by the pharmacokinetic properties of  $abi$  and  $ci^*$ .

Straightforward variations of the  $abi$  and  $ci^*$  designs that utilize IC and the properties of antiinhibitors would allow selective targeting of any one of the eight cell types that differ by the presence or absence of three macromolecular targets. Moreover, there is no formal limit on the total number of negative and/or positive targets that can be simultaneously sensed by regimens that employ  $abi$ - and  $ci^*$ -type compounds bearing multiple interference moieties. Note that the cell type selectivity of regimens such as  $abi + ci^*$  (Fig. 4) is analogous to the outcomes of the Boolean operations "and," "or," and combinations thereof.

**Stoichiometries, Affinities, and Pharmacokinetics.** The selectivity of the proposed compounds results from mutually exclusive, competing interactions between individual moieties of these compounds and their macromolecular targets (Figs. 2–4); hence, the importance of the targets' intracellular concentrations, relative to each other and the enzyme  $I$ , which is inhibited by an effector moiety of these drugs. The choice of  $I$  is not confined to essential enzymes. The target  $I$  could be, for instance, a DNA-binding repressor of terminal differentiation, a repressor of apoptosis, or, in the example of a noncytotoxic therapy, a neurotransmitter-inactivating enzyme. In other words, the choice of  $I$  is determined by the intended effect of the (unsequestered) compound  $abi$ .

The sequestration of  $abi$ - and  $ci^*$ -type compounds by their macromolecular ligands  $A$ ,  $B$ , and  $C$  serves to prevent their binding to the enzyme  $I$  (Figs. 3 and 4). Therefore, in schemes of the type considered above, the molar concentration of  $I$  should be significantly (if possible, considerably) lower than the molar concentrations of  $A$ ,  $B$ , and  $C$ . In addition, the concentration of  $ci^*$  in a  $ci^* + abi$  regimen should significantly exceed that of  $abi$ , because  $ci^*$  is the sole obstacle to the inhibition of enzyme  $I$  by  $abi$  in the  $A^- B^- C^-$  cells (Fig. 4). It is assumed, furthermore, that the total intracellular concentrations of  $abi$  and  $ci^*$ , bound and unbound, would remain significantly below the concentrations of the interference targets  $A$ ,  $B$ , and  $C$ . The affinities of  $A$ ,  $B$ ,  $C$ , and  $I$  for the

Cell type	Major complexes	Minor complexes	Toxicity of 
$A^+ B^+ C^+$			Low
$A^- B^+ C^+$			Low
$A^+ B^- C^+$			Low
$A^+ B^+ C^-$			Low
$A^+ B^- C^-$			Low
$A^- B^- C^+$			High
$A^- B^+ C^-$			Low
$A^- B^- C^-$			Low

As a result, a larger fraction of the inhibitor-containing **abi** molecules would be available for the interaction with **I**, resulting in the selective toxicity of **abi** to  $A^- B^- C^+$  cells. The selectivity pattern of **abi** requires that certain pharmacokinetic conditions are met as well (see the main text). Note that the physiological effects and the uses of **abi**-type compounds are not confined to cytotoxic regimens.

respective moieties of the compounds **abi** and **ci\***, and also the targets' intracellular locations are among the independent parameters that can be varied in designing these compounds.

Yet another, and major, constraint on the nature and pharmacokinetics of multitarget drugs stems from the fact that the selectivity patterns described above (Figs. 2 and 4) may not be observed under equilibrium conditions, where the influx of a drug into cells equals its outflux. To illustrate one clear difficulty, let us oversimplify and suppose that an **abi**-type compound is metabolically inert, in addition to being capable of crossing the plasma membranes and other lipid bilayers. If **abi** is initially outside the cells, and if the extracellular pool of **abi** (e.g., in the blood plasma) is large enough, it can be shown that the subsequently reached equilibrium state would be characterized by equal concentrations of the free **abi** in both the  $A^+ B^+$  and  $A^- B^-$  cells, thereby resulting in the equal occupancies of the enzyme **I** by **abi** in these cells, contrary to the pattern illustrated in Fig. 2.

By contrast, the selectivity patterns of Figs. 2 and 4 would be observed during the initial influx of drugs into cells. Thus, one requirement for the multitarget selectivity of **abi**- and **ci\***-type regimens is the avoidance of equilibrium states such as the one described above. This and related considerations indicate that despite the logical simplicity of the proposed designs, their implementation will have to address pharmacokinetic problems that do not necessarily arise with single-target drugs.

**Selection of Targets and Construction of Multitarget Drugs.** Appropriate macromolecular targets of the A-C class (Figs. 3 and 4) are suggested by the protein composition of the tumor cells to

FIG. 4. Combinatorial (positive/negative) selectivity of a regimen that utilizes interference/codominance and antiinhibitor. The diagram tabulates the relative toxicities of the compound **abi** in the setting of a two-compound treatment of cells that either lack or contain the macromolecular targets **A**, **B**, and **C**. The inhibitor-containing compound **abi** is described in the main text and Figs. 1, 2, and 3C. The antiinhibitor-containing compound **ci\*** is described in the main text and Fig. 3A and B are negative targets, in that they reduce, through the binding to the moieties **a** and **b** of **abi**, the inhibition of an essential enzyme **I** by **abi**. **C** is a positive target, in that it reduces, through the binding to the **c** moiety of **ci\***, the binding of **ci\*** to enzyme **I**. This results in a larger fraction of enzyme **I** available for the inhibition by **abi**. It is assumed that the concentrations of **A**, **B**, and **C** in cells that contain them significantly exceed the concentration of **I** (see the main text). In all of the cell types except  $A^- B^- C^+$ , the enzyme **I** would be at most partially inhibited by the **i** moiety of **abi**, because of the competing interactions of **abi** with **A** and/or **B**, and also because in  $C^-$  cells a fraction of enzyme **I** would be protected from the inhibition by **abi** through the interaction of **I** with the antiinhibitor moiety **i\*** of **ci\***. By contrast, in  $A^- B^- C^+$  cells, the antiinhibitor-containing **ci\*** would be sequestered by **C**, whereas **abi** would not be sequestered by **A** or **B**, which are absent from these cells.

be eliminated. The choice of an essential intracellular enzyme **I** (Figs. 2-4) is determined by the presence of **I** at least in tumor cells, its physiological concentration, and the feasibility of an efficacious inhibitor of **I**. Among potentially suitable enzymes for which cell-penetrating inhibitors already exist is dihydrofolate reductase. Its high-affinity inhibitors include methotrexate, which enters cells through carrier-mediated pathways, and the more lipophilic trimetrexate, which can enter cells by diffusing through lipid bilayers (18).

Each of the modules in the **abi**- and **ci\***-type compounds (Figs. 1-4) would bind its macromolecular target in the absence of the other modules. Therefore, the cytotoxic **i**-type modules of IC-based compounds would be similar to the stand-alone cytotoxic drugs of today. By contrast, the interference modules of these compounds, i.e., their **a**-, **b**-, and **c**-type moieties (Figs. 1-3), are supposed to bind to their macromolecular targets but preferably not impair them functionally. This specification of a competition module simplifies its design in comparison to that of inhibitors, because many sites on the target's surface, and not just the active site, would be acceptable.

Antiinhibitors (Fig. 3) are a new class of physiologically active compounds. The opportunities and problems of their design are similar to those for the interference moieties **a-c** (Figs. 1-3), but there are two other difficulties as well. First, the antiinhibitor moiety **i\*** must bind in the vicinity of, but not at, the active site of enzyme **I**. Second, the moiety **i\*** must also bear chemical groups (a bump) whose function is to preclude, through steric hindrance, the binding of the inhibitor moiety **i** to the enzyme **I** (Fig. 3).

A substrate-binding cleft is not the only indentation in a folded protein molecule. The other clefts tend to be smaller, but they are present as well (19), and some of them may be located next to the enzyme's active site (Fig. 3). In addition, even relatively flat molecular surfaces can be, in principle, the sites of high-affinity interactions with small ligands (20). These optimistic comments notwithstanding, the development of antiinhibitors is certain to be a complex undertaking. As discussed above, the pharmacokinetic aspects of the proposed designs are also complex. Yet simplicity is good only if it works. Single-target anticancer drugs remain unsatisfactory, in spite of decades of immense effort. It may therefore be wise to attempt a more complex but also more effective solution.

A recent advance in drug design, termed SAR by NMR (structure-activity relationships by nuclear magnetic resonance), provides an especially promising route to constructing ligands for specific regions of a protein molecule (21). In this approach, a library of small molecules is screened for binding to an <sup>15</sup>N-labeled protein by using NMR, which can detect weak interactions, and in addition assigns them to specific nitrogens of a protein, thereby identifying the site of binding. Finding two small compounds that bind to adjacent patches of the target protein molecule and covalently linking these compounds produces a higher-affinity ligand. This powerful strategy (21), which already yielded tightly binding ligands of specific proteins, may prove sufficient for constructing the **a-c** competition modules and the **i\*** antiinhibitor modules of the proposed designs (Figs. 1–4).

**Noncytotoxic Multitarget Drugs.** Many useful drugs are the inhibitors of intracellular enzymes that are not essential for cell viability (2). The problem of insufficient selectivity is relevant to these drugs as well. For example, even an exquisitely specific inhibitor of a clinically relevant enzyme is likely to have significant side effects, because the target enzyme is present, in most cases, not only in the cells where its inhibition is clinically beneficial but also in the cells where its inhibition is physiologically inappropriate. The logic of **abi**-type inhibitors (Figs. 1 and 2) and **ci\***-type antiinhibitors (Figs. 3 and 4) is applicable in these settings, because an informed choice of the competition moieties **a**, **b**, and **c** would sharpen up the cell selectivity of the inhibitor moiety **i** in the way described above for cytotoxic drugs (Fig. 4), resulting in the inhibition of the (nonessential) enzyme **I** in a predetermined subset of the enzyme-containing cells. Note that the same considerations apply to extracellular settings as well. The examples above are but a glimpse of the drug-engineering vistas that are opened up by the IC and antieffector concepts. At the same time, there are significant pharmacokinetic constraints on the properties of the proposed drugs, as discussed above. These constraints are likely to complicate the implementation of the IC/antieffector strategies.

**The Problem of Drug Resistance.** With small anticancer drugs that are in use today, the macromolecular target of a drug serves two distinct functions. First, the target is a cell-selectivity determinant that may bias the treatment against tumor cells. Second, the target is also a device whose inhibition by the drug brings about the desired effect, e.g., cell death. Consequently, when drug-resistant tumor cells, selected by a drug treatment, present themselves as a cancer recurrence, the necessity of employing another therapeutic agent (if such an option exists) robs the physician of whatever cell-selectivity advantage there was with the earlier drug.

The situation is qualitatively different with IC-based compounds. Suppose that a treatment that included the drug **abi** (Figs. 1 and 2) results in the appearance of **abi**-resistant tumor cells that contain, for example, an altered or overproduced enzyme **I**. If so, replacement of the **i** moiety by another small cytotoxic moiety, specific for another essential enzyme, would retain the cell selectivity of the new **ab**-containing drug. Thus,

one advantage of modularity inherent in the designs of IC/antiinhibitor-based compounds (Figs. 1–4) lies in the separation of the effector aspect of a drug from its selectivity aspect. As a result, once an efficacious arrangement of the selectivity modules in **abi**- or **ci\***-type compounds has been identified, it can be reutilized in drugs bearing effector moieties other than **i** and **i\***.

**Concluding Remarks.** The above considerations are based on the existing understanding of single-target drugs and on the notion of steric hindrance. By introducing the new concept of antieffectors and a modification of the previously proposed idea of codominant interference (14), we can now attempt the construction of small modular compounds that possess a multitarget, combinatorial selectivity (Fig. 4). The IC/antieffector strategies are not confined to cytotoxic therapies and are relevant, in principle, to all pharmacological settings. As indicated above, one expected difficulty in implementing these strategies stems from significant pharmacokinetic constraints that do not necessarily arise with single-target drugs.

This work was motivated by the premise that the confinement of anticancer drug research and development to single-target compounds will prove insufficient for the task at hand, because even the informed choices of targets for such drugs may not define unambiguously enough the cell type to be eliminated. The remedy, described above, is to aim for drugs that possess qualitatively different selectivity—multitarget and combinatorial. If this view is correct, the future ascent of multitarget drugs may transform not only the treatment of cancer but also approaches in other settings where the killing or modification of undesirable cells or organelles is carried out in the presence of nearly identical cells or organelles that must be spared. These applications of multitarget drugs encompass more discriminating antiviral and antifungal therapies, as well as the selective killing of activated lymphocytes in autoimmune diseases and the selective elimination of damaged mitochondria in aging cells (14, 15). In yet another class of applications, a noncytotoxic multitarget drug would be used to inhibit a clinically relevant nonessential enzyme in a specific subset of the enzyme-containing cells, thereby retaining the benefits of inhibition while reducing its side effects.

I thank L. Peck, G. Turner, D. Anderson, A. Rich, S. Mayo, E. Berezutskaya, and A. Kashina for comments on the manuscript. Studies in my laboratory are supported by grants from the National Institutes of Health.

1. Gilman, A. G., Rall, T. W., Nies, A. S. & Taylor, P. (1990) *The Pharmacological Basis of Therapeutics* (Pergamon, New York).
2. Munson, P. L., Mueller, R. A. & Breese, G. R. (1996) *Principles of Pharmacology* (Chapman & Hall, New York).
3. Waldman, T., Zhang, Y., Dillehay, L., Yu, J. K., Vogelstein, B. & Williams, J. (1997) *Nat. Med.* **3**, 1034–1036.
4. Folkman, J. (1985) *Adv. Cancer Res.* **43**, 175–203.
5. Thrush, G. R., Lark, L. R., Clinchy, B. C. & Vitetta, E. S. (1996) *Annu. Rev. Immunol.* **14**, 49–71.
6. Rosenberg, S. A. (1996) *Annu. Rev. Med.* **47**, 481–491.
7. da Costa, L. T., Jen, J., He, T.-C., Chan, T. A., Kinzler, K. W. & Vogelstein, B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4192–4196.
8. Bischoff, J. R., Kirn, D. H., Williams, A., Heise, C., Horn, S., Muna, M., Ng, L., Nye, J. A., Sampson-Johannes, A., Fattacy, A. & McCormick, F. (1996) *Science* **274**, 373–376.
9. Bishop, J. M. (1995) *Genes Dev.* **9**, 1309–1315.
10. Kinzler, K. W. & Vogelstein, B. (1996) *Cell* **87**, 159–170.
11. Weinberg, R. A. (1995) *Cell* **81**, 323–330.
12. Knudson, A. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10914–10921.
13. Hunter, T. (1997) *Cell* **88**, 333–346.
14. Varshavsky, A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3663–3667.
15. Varshavsky, A. (1996) *Cold Spring Harbor Symp. Quant. Biol.* **60**, 461–478.
16. Bernstein, K. E., Welsh, S. L. & Inman, J. K. (1990) *Biochem. Biophys. Res. Commun.* **167**, 310–316.
17. Crabtree, G. R. & Schreiber, S. L. (1996) *Trends Biochem. Sci.* **21**, 418–422.
18. Takimoto, C. H. & Allegra, C. J. (1995) *Oncology* **9**, 649–659.
19. Laskowski, R. A., Luscombe, N. M., Swindells, M. B. & Thornton, J. M. (1996) *Protein Sci.* **5**, 2438–2452.
20. Mattos, C. & Ringo, D. (1996) *Nat. Biotech.* **14**, 595–599.
21. Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1997) *Science* **278**, 497–499.

# The mouse and human genes encoding the recognition component of the N-end rule pathway

(ubiquitin/proteolysis/E3/N-recognin/Ubr1)

YONG TAE KWON\*, YUVAL REISS†, VICTOR A. FRIED‡, AVRAM HERSHKO§, JEONG KYO YOON\*, DAVID K. GONDA¶, PITCHAI SANGAN¶, NEAL G. COPELAND||, NANCY A. JENKINS||, AND ALEXANDER VARSHAVSKY\*,\*\*

\*Division of Biology, California Institute of Technology, Pasadena, CA 91125; †Department of Biochemistry, Tel Aviv University, Tel Aviv 69978, Israel;

‡Department of Cell Biology and Anatomy, New York Medical College, Valhalla, NY 10595; §Unit of Biochemistry, Faculty of Medicine, Technion, Haifa 31096, Israel; ¶Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520-8024; and ||Mammalian Genetics Laboratory, Advanced BioScience Laboratories-Basic Research Program, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD 21702

Contributed by Alexander Varshavsky, May 5, 1998

**ABSTRACT** The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The N-end rule pathway is one proteolytic pathway of the ubiquitin system. The recognition component of this pathway, called N-recognin or E3, binds to a destabilizing N-terminal residue of a substrate protein and participates in the formation of a substrate-linked multiubiquitin chain. We report the cloning of the mouse and human *Ubr1* cDNAs and genes that encode a mammalian N-recognin called E3 $\alpha$ . Mouse UBR1p (E3 $\alpha$ ) is a 1,757-residue (200-kDa) protein that contains regions of sequence similarity to the 225-kDa Ubr1p of the yeast *Saccharomyces cerevisiae*. Mouse and human UBR1p have apparent homologs in other eukaryotes as well, thus defining a distinct family of proteins, the UBR family. The residues essential for substrate recognition by the yeast Ubr1p are conserved in the mouse UBR1p. The regions of similarity among the UBR family members include a putative zinc finger and RING-H2 finger, another zinc-binding domain. *Ubr1* is located in the middle of mouse chromosome 2 and in the syntenic 15q15-q21.1 region of human chromosome 15. Mouse *Ubr1* spans  $\approx$ 120 kilobases of genomic DNA and contains  $\approx$ 50 exons. *Ubr1* is ubiquitously expressed in adults, with skeletal muscle and heart being the sites of highest expression. In mouse embryos, the *Ubr1* expression is highest in the branchial arches and in the tail and limb buds. The cloning of *Ubr1* makes possible the construction of *Ubr1*-lacking mouse strains, a prerequisite for the functional understanding of the mammalian N-end rule pathway.

A number of regulatory circuits involve metabolically unstable proteins. Short *in vivo* half-lives are also characteristic of damaged or otherwise abnormal proteins (1–4). Features of proteins that confer metabolic instability are called degradation signals, or degrons. The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (5, 6). The set of amino acid residues that are destabilizing in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar, but distinct, versions of the N-end rule pathway are present in all organisms examined, from mammals to fungi and bacteria (6–8).

In eukaryotes, the N-degron comprises two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (8). The Lys residue is the site of formation of a multiubiquitin chain (9). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/957898-6\$2.00/0  
PNAS is available online at <http://www.pnas.org>.

conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (1, 3, 4, 10). In most of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (11).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their enzymatic deamidation into the secondary destabilizing N-terminal residues Asp and Glu (12). The destabilizing activity of N-terminal Asp and Glu requires their enzymatic conjugation to Arg, one of the primary destabilizing residues (6). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3), the recognition component of the N-end rule pathway (13). In *S. cerevisiae*, N-recognin is a 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues—Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His. N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (6).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae* (through degradation of Cup9p, a transcriptional repressor of the peptide transporter Ptr2p); a role in controlling the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae*; the degradation of Gpa1p, a G $\alpha$  protein of *S. cerevisiae*; and the degradation of alphavirus RNA polymerases in virus-infected metazoan cells (6, 14).

The mammalian counterpart of the yeast *UBR1*-encoded N-recognin (E3) was characterized biochemically in extracts from rabbit reticulocytes (15–17). Rabbit E3 $\alpha$  was shown to be spe-

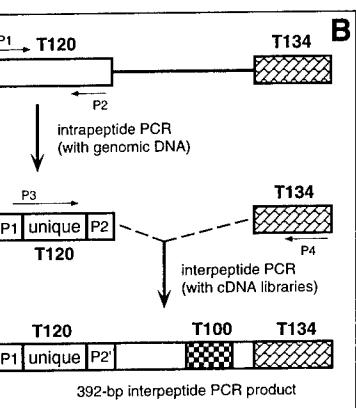
Abbreviations: Ub, ubiquitin; kb, kilobase; id., identity; si., similarity; BAC, bacterial artificial chromosome; FISH, fluorescence *in situ* hybridization; en, embryonic day.

Data deposition: Nucleotide sequences reported in this work have been deposited in the GenBank database [accession nos. AF061555 (mouse *Ubr1* cDNA) and AF061556 (human *UBR1* cDNA)].

\*\*To whom reprint requests should be addressed at: Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. e-mail: avarsh@cco.caltech.edu.

††The names of mouse genes are in italics, with the first letter uppercase. The names of human and *S. cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. The names of *S. cerevisiae* proteins are Roman, with the first letter uppercase and an extra lowercase "p" at the end. The names of the corresponding mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention (33), to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.

A Peptide	Determined sequence	Position
T70B	YMSYSYDFR	324-332
T70A	FNFQGYSQDK	459-468
T98	GILISKPTISIER	480-493
T120	QVGQHIEVDPDWEAAIAYQMLK	521-543
T100 (PEP3)	VSEDLVSLNPLSYSK	600-615
T134	FVPFEDFHIEVLVEYPLR	634-651
T74B	VPOEFNVTXEVTM	755-769
T76 (PEP2)	NLPENNETGLENVINK	789-806
T92	FNMYFYHYSK	831-840
T96 (PEP1)	APEEEWDFYHK	932-944
T122	GETLDPLMFDPLAYGTY	1140-1158
T74A	YFEAVQLSSQQR	1174-1185
T132	VDLFOLSEGEYL	1188-1199
T140	QETNQMLFGFN	1743-1753



the rabbit *Ubr1* cDNA was used to isolate, using PCR and a λgt11 mouse liver cDNA library, the corresponding mouse *Ubr1* cDNA fragment. This fragment then was used to screen the same cDNA library, yielding a 2.4-kb fragment of the mouse *Ubr1* cDNA that encoded several of the peptide-derived sequences of the rabbit UBR1p. The encoded sequence was also significantly similar to that of the N-terminal region of *S. cerevisiae* Ubr1p (13) and contained the putative start (ATG) codon of the mouse *Ubr1* ORF. To isolate the rest of the 5' region of the *Ubr1* cDNA, 5'-rapid amplification of cDNA ends (RACE)-PCR (20) was performed with poly(A)<sup>+</sup> RNA from mouse L cells and a primer from the 2.4-kb DNA fragment. 3'-RACE-PCR (20) was used to amplify a downstream region of *Ubr1* cDNA. The resulting DNA fragment (nucleotides 2,470–3,467) then was used to screen a λgt10 mouse cDNA library from MEL-C19 cells. Five overlapping cDNA isolates (MR16, MR17, MR19, MR20, and MR23) that together spanned the entire *Ubr1* cDNA were mapped and subcloned into Bluescript II SK<sup>+</sup> (Stratagene), yielding the plasmid MR26, which contained the entire ORF of *Ubr1*. The ORF region of *Ubr1* cDNA was sequenced on both strands at least twice, using independently derived cDNA clones.

cifically required for the Ub-dependent degradation of proteins bearing either type 1 (basic) or type 2 (bulky hydrophobic) destabilizing N-terminal residues (7, 15, 16).

We began dissection of the mouse N-end rule pathway by isolating the *Ntan1* gene, which encodes the asparagine-specific N-terminal amidase (18, 19), a component of the mammalian N-end rule pathway, and by constructing mouse strains that lack *Ntan1* (Y.T.K. and A.V., unpublished data). Herein, we describe the cloning and characterization of the mouse and human cDNAs and genes<sup>††</sup> that encode UBR1p (E3 $\alpha$ ), a homolog of yeast Ubr1p and the main recognition component of the N-end rule pathway.

## MATERIALS AND METHODS

**Isolation and Partial Sequencing of Mammalian E3 $\alpha$  (UBR1p).** Rabbit E3 $\alpha$  was purified from reticulocyte extracts by using affinity chromatography with immobilized protein substrates of UBR1p and elution with dipeptides bearing destabilizing N-terminal residues (16). The resulting preparation was

fractionated by SDS/PAGE. The band of ≈180-kDa E3 $\alpha$  was excised and subjected to digestion with trypsin. Amino acid sequences were determined for 14 peptides of rabbit UBR1p (Fig. 1A) by using standard methods (20).

**Isolation of the Full-Length Mouse *Ubr1* cDNA.** A strategy that included the intrapeptide-interpeptide PCR (21) was used (see the legend to Fig. 1).

**Isolation of a Partial Human *UBR1* cDNA.** Poly(A)<sup>+</sup> RNA from human 293 cells was subjected to reverse transcription-PCR, using sets of primers corresponding to sequences of the mouse *Ubr1* cDNA. One of the reactions yielded a 1.0-kilobase (kb) fragment that encompassed a region of the human *UBR1* cDNA (Fig. 2).

**Mouse and Human Genomic *Ubr1* Fragments.** A library of mouse genomic DNA fragments (strain SVJ) in bacterial artificial chromosome (BAC) (22) vector (Genome Systems, St. Louis) was used, as described in the legend to Fig. 2.

**Northern, Southern, and Whole-Mount *In Situ* Hybridizations.** Mouse and human multiple-tissue Northern blots

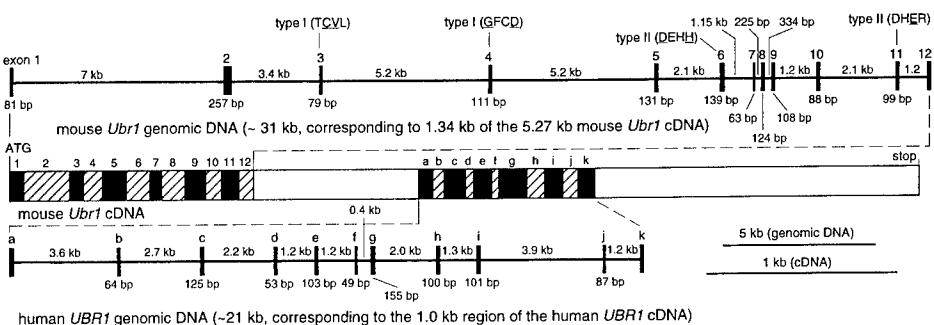


diagram is a ≈21-kb fragment of the human *UBR1* gene, corresponding to 1.0 kb of the indicated region of the human *UBR1* cDNA (nucleotides 2,218–3,227 of the mouse *Ubr1* cDNA sequence). The mouse and human *Ubr1* exons are denoted, respectively, by numbers and letters. Also indicated are the exon locations of some of the type I and type 2 substrate-binding sites of N-recognin (the essential amino acid residues are underlined) (A. Webster, M. Ghislain, and A.V., unpublished data; see the main text). Not shown are the 114-bp 5'-untranslated region (UTR) and the 1,010 bp 3'-UTR of the mouse *Ubr1* cDNA. To isolate mouse *Ubr1*, a library of mouse genomic DNA fragments in a BAC vector (see Materials and Methods) was screened with a fragment of the mouse *Ubr1* cDNA (nucleotides 105–1,333) as a probe, yielding seven BAC clones, of which BAC3 and BAC4 contained the entire *Ubr1* gene. The exon/intron organization of the first 31 kb (≈1/4) of the mouse *Ubr1* gene was determined by using exon-specific PCR primers to produce ≈40 genomic DNA fragments of the BAC3 insert that ranged in size from 1.3 to 18 kb. Regions encompassing the exon/intron junctions then were sequenced by using intron-specific primers. Fragments of the human genomic *UBR1* DNA were isolated by using primers derived from the 1.0-kb fragment of the human *UBR1* cDNA, the Expand High Fidelity PCR System (Roche Molecular Biochemicals, Indianapolis, IN), and genomic DNA from human 293 cells. The resulting four fragments were subcloned into pCR2.1 (Invitrogen), yielding the plasmids HR8, HR6–4, HR2–25, and HR7–2, whose partially overlapping inserts encompassed ≈21 kb of the human *UBR1* gene. Partial sequencing of the mouse and human genomic *Ubr1* fragments (≈20 kb of sequenced DNA) included all of the exon/intron junctions in these regions of *Ubr1*.

FIG. 2. The mouse and human *Ubr1* cDNAs and genes. Thick horizontal lines represent genomic DNA. The upper one is a ≈31-kb fragment of the mouse *Ubr1* gene that corresponds to a 1.34-kb region of mouse *Ubr1* cDNA (nucleotides 115–1,454). Vertical rectangles represent exons. Their lengths, and the lengths of the introns, are indicated, respectively, below and above the horizontal line. In a composite diagram of the *Ubr1* cDNA, the exons are depicted as alternatively shaded rectangles. For exon 1, only its translated region is indicated. Shown below the cDNA

is the 5 kb (genomic DNA) and 1 kb (cDNA).



Region I (Cys/His region)													
	C	C	C	C	C	C	H	H	C	C	C	C	C
Mm-UBR1	CCKVFKS	ETTY	SERDCA	IDPTCVLCMDCF	QSSVHK	NHRY	--KMHTSTG	GGF	CDCCG	TEAWKTGF	C		
Ce-UBR1	CGHVFKNG	ELTY	TCLDCA	TDGTCV	MCLQCF	EVSIHK	SHKN	--KMHSSSG	SGY	CDCCG	ADAWT	EYAC	
Sc-UBR1	CGRKF	K1GEPLY	RCHECG	CDDICVLC	TCENPKD	DHVNVH	HVCTDICTEFT	SGT	CDCCG	EEAWNSPL	HC		
Ca-UBR1	OAKP	OPDPVYR	RODECG	FDDTCVLC	CASECN	KEHDHN	NNTVYT	SSGTG	GG	ICDCCG	P		
K1-UBR1	CGAKF	VRGEPIYR	KECS	FFDT	CVLC	CVNCF	NPKD	DHVGVH	HYVTS	ICDCG	KEA	WNHELN	C
Ce-UBR2	ONEIWENDA	VAFRCNC	TA	LTPCM	SLCEDCF	ESN	GAGHDY	--TRFFSRE	GGQAC	DCCNODV	VIKEOGNC		
Sc-UBR2	QTRLCP	FPSETIY	YCF	TCSTNPL	YE	IGELCF	DKEKHVNHS	VAKVV	MRPE	GRICH	CGGPFAFN	NDPSDA	

Region IV (RING-H2 finger)																
	C	C	C	H	H	C	C	C	C	C	C	C	C			
Mm-UBR1	T	CIL	COEE	-53-	SCGH	HVM	A	WCWOK	Y	FEAVQ	S	-15-	EYL	CPLO		
Ce-UBR1	T	CIL	COED	-59-	TC	CSH	S	HYEC	YRSLAEAN	RSR	-21-	EYOC	CPLO	KRL		
Sc-UBR1	T	CAL	CQDS	-67-	SCN	HH	I	HNC	FKR	YOKKRF	-2-	AF	ICPL	QOTFSN	TLPL	
K1-UBR1	H	CSL	CHDD	-70-	SCN	H	A	H	YRCFKH	YDKKRY	-22-	LF	ICPL	QTYCNS	VIP	
Ce-UBR2	E	CP	TCG	GEQ	-85-	TCG	HAH	I	ACFNA	YASLVG	-23-	EVG	CP	MCYRTV	VNVV	
Sc-UBR2	D	CVF	CKMP	-68-	ACG	GHG	SH	T	KCL	SGHMK	SIRGI	-15-	LI	YCPV	CNSL	SNSFLPK

(FISH) with mitotic chromosomes from human lymphocytes (25). The probe was a mixture of the HR8, HR6–4, HR2–25, and HR7–2 plasmids, labeled with biotin using biotinylated dATP and the BioNick labeling kit (Life Technologies, Grand Island, NY), and detected by using fluorescein isothiocyanate-avidin.

## RESULTS AND DISCUSSION

**Isolation of the Mouse *Ubr1* cDNA.** Tryptic peptides of the purified rabbit E3 $\alpha$  (UBR1p) (16) were isolated and sequenced (see Materials and Methods), yielding 14 short regions of E3 $\alpha$  (Fig. 1A). These regions lacked significant similarities to the deduced sequence of *S. cerevisiae* Ubr1p (13). We used intrapeptide PCR (21) to identify a unique (nondegenerate) sequence of the rabbit *Ubr1* cDNA. This method allows amplification of a short unique DNA sequence by using two degenerate PCR primers (derived by reverse translation) that flank this sequence and correspond to the outermost regions of a single peptide (Fig. 1B). Several intrapeptide nucleotide sequences were obtained this way (Fig. 1). These sequences, together with those of the original degenerate primers, then were used to amplify a 392-bp fragment of the rabbit *Ubr1* cDNA, using interpeptide PCR (Fig. 1). This fragment encoded peptides T120 and T134 at either end and peptide T100 in the middle (Fig. 1B). A homologous 392-bp fragment of the mouse *Ubr1* cDNA then was amplified by using the same method (Fig. 1B). The rabbit and mouse 392-bp *Ubr1* cDNA fragments were 88% and 89% identical at the nucleotide and amino acid sequence levels, respectively, but lacked significant similarities to *S. cerevisiae* *UBR1* (data not shown).

The 392-bp mouse *Ubr1* cDNA fragment then was used, in conjunction with standard cDNA library screening and rapid amplification of cDNA ends-PCR (20), to isolate multiple *Ubr1* cDNA fragments, and to assemble them into a 5,271-bp ORF encoding a 1,757-residue protein (pI of 6.0), whose size, 200 kDa, was close to the estimated size of the isolated rabbit UBR1p (E3 $\alpha$ ), ≈180 kDa (16) (Figs. 2 and 3). The inferred ATG start codon (Fig. 2), within the sequence CTTAACATGGCG, is preceded by two in-frame stop codons, at positions -48 and -93, and is located in a favorable Kozak context (26), with A and G at positions -3 and +4, respectively. There are two more ATGs, five and 11 codons downstream of the inferred one. These alternative start codons are in a favorable Kozak context as well.

**Cloning and Partial Characterization of the Mouse and Human *Ubr1* Genes.** A fragment of the mouse *Ubr1* cDNA was used to isolate a ≈120-kb mouse *Ubr1* genomic DNA clone, carried in a BAC vector (22). We determined the exon/intron organization and restriction map of the ≈31-kb region of *Ubr1* that corresponded to the 1,340-bp 5'-region of the mouse *Ubr1* cDNA (nucleotides 105–1,333) (Fig. 2). The lengths of the 12 exons in this region of mouse *Ubr1* range from 63 to 257 bp (Fig. 2).

The nucleotide and deduced amino acid sequences of the 1.0-kb human *UBR1* cDNA fragment (see Materials and Methods), located approximately in the middle of *UBR1* cDNA (nucleotides 2,218–3,227 of the mouse *Ubr1* cDNA) (Fig. 2), were, respectively, 91% and 94% identical to the corresponding mouse *Ubr1* cDNA and UBR1p sequences. Overlapping genomic

Fig. 4. Two Cys/His domains of the *UBR* protein family. Comparison of the putative zinc finger (region I) and RING-H2 finger (region IV) with the corresponding sequences from the other species in Fig. 3, and also with *C. albicans* Ubr1p (Ca-UBR1), *C. elegans* UBR2p (Ce-UBR2), and *S. cerevisiae* Ubr2p (Sc-UBR2). Numbers indicate the lengths of gaps. The conserved Cys and His residues are indicated.

DNA fragments of human *UBR1* that, together, encompassed a ≈21-kb region of the human *UBR1* gene and corresponded to the 1.0-kb fragment of the human *UBR1* cDNA (Fig. 2), were isolated from human DNA by using cDNA-derived primers and PCR. Partial sequencing showed that this ≈21-kb region of human *UBR1* contained 11 exons whose length ranged from 49 to 155 bp, a distribution of exon lengths similar to that in a different region of mouse *Ubr1* (Fig. 2). All of the sequenced exon/intron junctions (≈23 exons), which encompassed a ≈52-kb region of the mouse and human *Ubr1*, contained the consensus GT and AG dinucleotides characteristic of the mammalian nuclear pre-mRNA splice sites (data not shown) (20). Extrapolating from these data on the mouse and human *Ubr1* genes and the corresponding regions of their cDNAs (Fig. 2), a mammalian *Ubr1* gene is expected to be ≈120 kb long and to contain ≈50 exons.

**The Mouse UBR1p Protein and its Homologs.** The low overall sequence similarity of mouse UBR1p (E3 $\alpha$ ) to Ubr1p of either *S. cerevisiae* [22% identity (id.), 48% similarity (si.)] or another budding yeast, *Kluyveromyces lactis* (21% id., 48% si.), belied the presence of five regions, denoted I–V, which were significantly similar between the mouse and yeast versions of UBR1p (Figs. 3 and 4). By contrast, the *Ntn1*-encoded asparagine-specific N-terminal amidase, the most upstream component of the mouse N-end rule pathway, lacks sequence similarities to its *S. cerevisiae* counterpart Nta1p (12, 18). Database searches identified other likely homologs of mouse UBR1p, in particular the 1,927-residue protein of the nematode *Caenorhabditis elegans* (GenBank accession no. U88308) (32% id., 53% si.; termed *Ce-Ubr1*); the 1,872-residue *S. cerevisiae* protein (GenBank accession no. Z73196) (21% id., 47% si.; termed *Sc-UBR2*; ref. 4); the 2,168-residue *C. elegans* protein (GenBank accession no. U40029) (21% id., 45% si.; termed *Ce-Ubr2*); and the 794-residue CER3p protein of the plant *Arabidopsis thaliana* (GenBank accession no. X95962) (26% id., 49% si.). CER3p is involved in wax biosynthesis in *A. thaliana* (27). In addition, a 147-residue sequence of the yeast *Candida albicans* (<http://alces.med.umn.edu/bin/genelist?LUBR1>) was similar to the N-terminal region of mouse UBR1p (Fig. 4).

The presence of high-similarity regions I–V among these deduced sequences (Figs. 3 and 4) suggested the existence of a distinct protein family, termed UBR. The 66-residue region I, near the N terminus of UBR1p, is a particularly clear UBR family-identifying region (e.g., 61% id., 75% si. between mouse and *C. elegans* UBR1p) (Figs. 3 and 4).

Recent genetic analyses of *S. cerevisiae* Ubr1p (N-recognin) have shown that the regions I–III contain residues essential for the recognition of N-end rule substrates by Ubr1p. In particular, Cys-145, Val-146, Gly-173, and Asp-176 of region I were identified as essential residues of the type 1 binding site of *S. cerevisiae* Ubr1p (A. Webster, M. Ghislain, and A.V., unpublished data). All four of these residues were conserved between the yeast, mouse, and *C. elegans* UBR1p (Figs. 3 and 4). Region I is present in all of the known UBR family members except CER3p of *A. thaliana*, which contains only regions IV and V (Fig. 4). Region I encompasses a Cys/His-rich domain, Cys-X<sub>12</sub>-Cys-X<sub>2</sub>-Cys-X<sub>5</sub>.

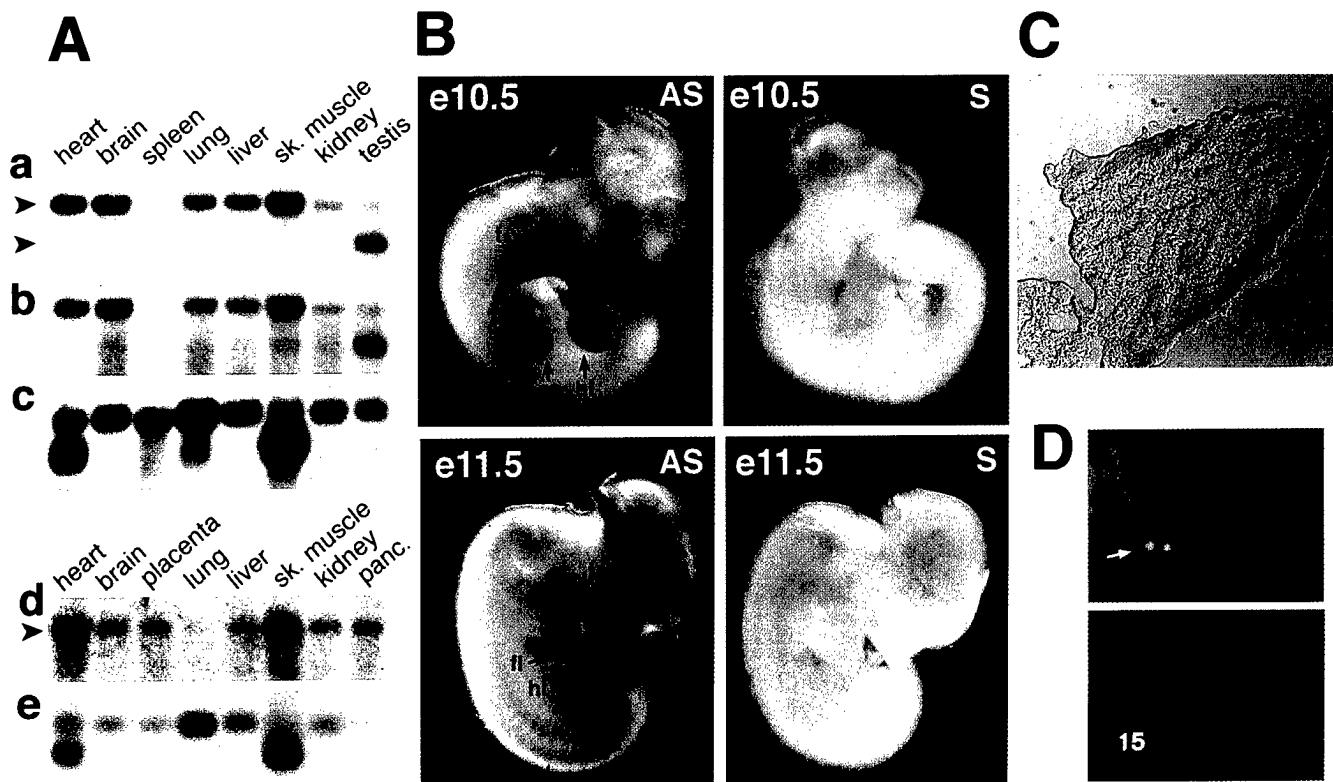


FIG. 5. Northern and *in situ* hybridizations with mouse and human *Ubr1*. (A) Membranes containing electrophoretically fractionated poly(A)<sup>+</sup> mRNA from different mouse (a–c) or human (d and e) tissues were hybridized with either a 2-kb 5'-proximal (nucleotides 116–2,124) mouse *Ubr1* cDNA fragment (a), its 0.64-kb 3'-proximal (nucleotides 4,749–5,388) fragment (b), a 1-kb human *UBR1* cDNA fragment (d), or the human  $\beta$ -actin cDNA fragment (c and e). The upper arrows in a and d indicate the ≈8-kb *Ubr1* transcript. The lower arrow in a indicates the ≈6-kb testis-specific *Ubr1* transcript. In the RNA sample from mouse spleen, the *Ubr1* transcript (but not the actin transcript) may have been degraded (a–c). (B) Expression of *Ubr1* in e10.5 and e11.5 mouse embryos. Whole-mount *in situ* hybridization was carried out with either antisense (AS) or sense (S, negative control) *Ubr1* cDNA probes (see Materials and Methods). The regions of high *Ubr1* expression are indicated by arrows (t, tail; fl, forelimb buds; hl, hindlimb buds). The branchial arches, where *Ubr1* is also highly expressed in e10.5 embryos (data not shown), are not visible in this e10.5 embryo. (C) Expression of *Ubr1* in the surface ectoderm of limb buds. Shown is a transverse section of a forelimb bud of an e10.5 embryo (se, surface ectoderm). (D) FISH analysis of human *UBR1*. (Upper) An example of the *UBR1*-specific FISH signal (arrow). (Lower) The same mitotic spread stained with 4'-6-diamino-2-phenylindole (DAPI) to visualize the chromosomes (see also Fig. 6).

Cys-X<sub>2</sub>-Cys-X<sub>2</sub>-Cys-X<sub>5</sub>-His-X<sub>2</sub>-His-X(12–14)-Cys-X<sub>1</sub>-Cys-X<sub>11</sub>-Cys (Figs. 3 and 4), which is distinct from the known consensus sequences of zinc fingers and other Cys/His-motifs. Residues Asp-318, His-321, and Glu-560 of *S. cerevisiae* Ubr1p, which have been identified as essential for the type 2 binding site of this N-recognin (A. Webster, M. Ghislain, and A.V., unpublished data), were found to be retained in region II (Asp-318 and His-321) and region III (Glu-560) of the mouse and *C. elegans* UBR1p (Fig. 3).

Region IV contains another Cys/His-rich domain of UBR1p, Cys-X<sub>2</sub>-Cys-loop 1-Cys-X<sub>1</sub>-His-X<sub>2</sub>-His-X<sub>2</sub>-Cys-loop 2-Cys-X<sub>2</sub>-Cys (Figs. 3 and 4), which is present in all of the UBR family members, and fits the consensus sequence of the RING-H2 finger, a subfamily of the previously defined RING motif (28). At least some of the RING-H2 sequences are sites of specific protein-protein interactions (28). Apc11p, a subunit of the Ub-protein ligase complex called the cyclosome (2) or the anaphase promoting complex, also contains a RING-H2 finger (29).

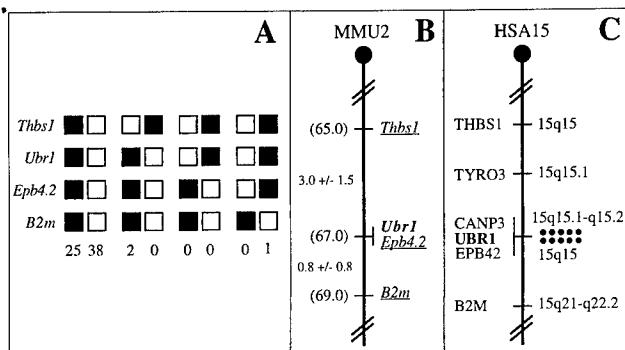
Another area of similarity (24–50% id., 46–70% si.) among the UBR family members is region V (Fig. 3 and data not shown). This region, 115 residues long in mouse UBR1p, near the protein's C terminus, is particularly similar between mouse and *C. elegans* UBR1p (50% id., 70% si.) (Fig. 3). Region V is located 4–14 residues from the UBR proteins' C termini, the exceptions being the *S. cerevisiae* and *K. lactis* Ubr1p, which bear, respectively, 132- and 159-residue tails of unknown function that are rich in the acidic Asp/Glu residues (36% and 33%) (Fig. 3). No significant similarities could be detected between mammalian UBR1p and other E3s (recognins) of the metazoan Ub system,

including E6AP (30) and subunits of the cyclosome/anaphase promoting complex, except for the presence of a RING-H2 finger domain in the latter (29). [Different E3 proteins of the Ub system recognize different degrons in protein substrates, thereby defining distinct Ub-dependent proteolytic pathways (1, 4).]

**Expression of Mouse and Human *Ubr1*.** The 5'- and 3'-proximal mouse cDNA probes yielded similar results, detecting a single ≈8-kb transcript in several tissues (Fig. 5Aa and Ab). In the testis, however, the ≈8-kb species of *Ubr1* mRNA was a minor one, the major species being ≈6 kb (Fig. 5Aa). The levels of either mouse or human *Ubr1* mRNA were highest in skeletal muscle and heart (Fig. 5A). The expression of mRNA encoding E2<sub>14K</sub>, one of the mouse Ub-conjugating (E2) enzymes and a likely component of the mouse N-end rule pathway (6), was also highest in skeletal muscle and heart (18).

The distinct *Ubr1* mRNA pattern in the testis (Fig. 5A) was reminiscent of the analogous expression pattern of *Ntan1* mRNA, which encodes the Asn-specific N-terminal amidase, another component of the mammalian N-end rule pathway. Specifically, the size of the major species of *Ntan1* mRNA was ≈1.4 kb in all of the examined mouse tissues except testis, where the major species was ≈1.1 kb (18). The ≈1.1-kb *Ntan1* transcript recently was found to hybridize only to the 3'-half (exons 6–10 but not exons 1–5) of the *Ntan1* ORF (Y.T.K. and A.V., unpublished data). The functional significance of the testis-specific *Ubr1* and *Ntan1* expression patterns remains to be understood.

We used whole-mount *in situ* hybridization to examine the expression of *Ubr1* during embryogenesis. In e9.5 (9.5 days old)



**FIG. 6.** Chromosomal locations of the mouse and human *Ubr1* genes. (A) Mouse *Ubr1* was mapped to the middle of mouse chromosome 2 by using interspecific (*M. musculus*-*M. spretus*) backcross analysis (18, 24). Shown are the segregation patterns of mouse *Ubr1* and the flanking genes in 66 backcross animals that were typed for all loci. For individual pairs of loci, more than 66 animals were typed. Each column represents the chromosome identified in the backcross progeny that was inherited from the [*M. musculus* C57BL/6J × *M. spretus*] F<sub>1</sub> parent. Filled and empty squares represent, respectively, C57BL/6J and *M. spretus* alleles. The numbers of offspring that inherited each type of chromosome 2 are listed below the columns. (B) A partial mouse chromosome 2 linkage map (MMU2), showing *Ubr1* in relation to the linked genes *Thbs1*, *Epb4.2*, and *B2m*, and also, on the left, the corresponding recombination distances between the loci, in centimorgans, and the map locations, in parentheses. (C) A partial human chromosome 15 linkage map (HSA15). Each dot on the right, in the 15q15-q21.1 region, corresponds to the actually observed *UBR1*-specific double-dot FISH signal detected on human chromosome 15 (see also Fig. 5D).

mouse embryos, the expression of *Ubr1* was highest in the branchial arches and in the buds of forelimbs and the tail (data not shown). In e10.5 embryos, the expression of *Ubr1* became high in the hindlimb buds as well (Fig. 5B). This pattern was maintained in the limb buds of e11.5 embryos (Fig. 5B). High expression of *Ubr1* in the limb buds was confined predominantly to the surface ectoderm (Fig. 5C). This pattern of *Ubr1* expression in embryos (Fig. 5B and C) is similar, if not identical, to that of *Ntan1*, which encodes asparagine-specific N-terminal amidase (18) (Y.T.K. and A.V., unpublished data), consistent with UBR1p and NTAN1p being components of the same pathway.

The enhanced expression of *Ubr1* in the embryonic limb buds (Fig. 5B and C) is interesting in view of the conjecture that the N-end rule pathway might be required for limb regeneration in amphibians (31). The injection of dipeptides bearing destabilizing N-terminal residues into the stumps of amputated forelimbs of the newt was observed to delay limb regeneration, whereas the injection of dipeptides bearing stabilizing N-terminal residues had no effect (31). Rigorous tests of this and other suggested functions of the metazoan N-end rule pathway (6) will require mouse strains that lack *Ubr1*.

**Chromosome Mapping of Mouse and Human *Ubr1*.** The chromosomal location of mouse *Ubr1* was determined by interspecific backcross analysis, using DNA derived from matings of [(C57BL/6J × *Mus spretus*)F<sub>1</sub> × C57BL/6J] mice (Fig. 6A and B) (18, 24). Mouse *Ubr1* is located in the central region of chromosome 2 and is linked to the *Thbs1*, *Epb4.2*, and *B2m* genes, the most likely gene order being centromere-*Thbs1*-*Ubr1*-*Epb4.2*-*B2m* (Fig. 6B and data not shown).

The chromosomal location of human *UBR1* was determined by using FISH (25), with human *UBR1* genomic DNA fragments as probes (Figs. 5D and 6C). This mapping placed *UBR1* at the 15q15-15q21.1 region of the human chromosome 15, an area syntenic with the independently mapped position of mouse *Ubr1* (Fig. 6). *Ubr1* is located in the regions of human chromosome 15 and mouse chromosome 2 that appear to be devoid of the previously mapped but uncloned mutations.

Mutations in the human gene *CANP3*, which encodes a subunit of calpain and is located very close, if not adjacent, to *UBR1*, have been shown to cause a myopathy called the limb-girdle muscular dystrophy (32).

**Concluding Remarks.** Isolation of the mouse and human *Ubr1* cDNAs and genes (Figs. 2–6) should enable functional understanding of the mammalian N-end rule pathway, in part through the construction and analysis of mouse strains that lack *Ubr1*. Recent searches in GenBank identified several mouse and human sequences in expressed sequence tag databases that exhibited significant similarity to the C-terminal region of mouse UBR1p. The cloning and characterization of the corresponding cDNAs have shown that there exist at least two distinct mouse (and human) genes, termed *Ubr2* and *Ubr3*, which encode proteins that are significantly similar to mouse UBR1p (Y.T.K. and A.V., unpublished data). Molecular and functional analyses of these *Ubr1* homologs are under way.

We are grateful to A. Webster and M. Ghislain for permission to cite their unpublished data. We thank members of the Varshavsky lab, especially I. V. Davydov, for helpful discussions, and L. Peck, G. Turner, H. Rao, A. Kashina, and F. Du for comments on the manuscript. Y.T.K. thanks B. Yu for sharing his Northern hybridization data on human β-actin mRNA. We gratefully acknowledge the sequencing of *K. lactis* *UBR1* by P. Waller. N.G.C. and N.A.J. thank D. J. Gilbert and D. B. Householder for excellent technical assistance. D.K.G. was a Scholar of the Leukemia Society of America. This study was supported by National Institutes of Health grants to A.V. (DK39520 and GM31530), V.A.F. (NS29542), and D.K.G. (GM45314), and by a grant to N.G.C. from the National Cancer Institute.

- Varshavsky, A. (1997) *Trends Biochem. Sci.* **22**, 383–387.
- Hershko, A. (1997) *Curr. Opin. Cell. Biol.* **9**, 788–799.
- Haas, A. J. & Siepmann, T. J. (1997) *FASEB J.* **11**, 1257–1268.
- Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30**, 405–439.
- Bachmair, A., Finley, D. & Varshavsky, A. (1986) *Science* **234**, 179–186.
- Varshavsky, A. (1997) *Genes Cells* **2**, 13–28.
- Gonda, D. K., Bachmair, A., Wünnig, I., Tobias, J. W., Lane, W. S. & Varshavsky, A. (1989) *J. Biol. Chem.* **264**, 16700–16712.
- Bachmair, A. & Varshavsky, A. (1989) *Cell* **56**, 1019–1032.
- Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989) *Science* **243**, 1576–1583.
- Pickart, C. M. (1997) *FASEB J.* **11**, 1055–1066.
- Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998) *Cell* **92**, 367–380.
- Baker, R. T. & Varshavsky, A. (1995) *J. Biol. Chem.* **270**, 12065–12074.
- Bartel, B., Wünnig, I. & Varshavsky, A. (1990) *EMBO J.* **9**, 3179–3189.
- Byrd, C., Turner, G. C. & Varshavsky, A. (1998) *EMBO J.* **17**, 269–277.
- Reiss, Y., Kain, D. & Hershko, A. (1988) *J. Biol. Chem.* **263**, 2693–2699.
- Reiss, Y. & Hershko, A. (1990) *J. Biol. Chem.* **265**, 3685–3690.
- Hershko, A. & Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61**, 761–807.
- Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. J. & Varshavsky, A. (1996) *J. Biol. Chem.* **271**, 28521–28532.
- Stewart, A. E., Arfin, S. M. & Bradshaw, R. A. (1995) *J. Biol. Chem.* **270**, 25–28.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1996) *Current Protocols in Molecular Biology* (Wiley Interscience, New York).
- Bredt, D. S., Hwang, P. M., Glatt, C. E., Lowenstein, C., Reed, R. R. & Snyder, S. H. (1991) *Nature (London)* **351**, 714–718.
- Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. I. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797.
- Conlon, R. A. & Rossant, J. (1992) *Development (Cambridge, U.K.)* **116**, 357–368.
- Copeland, N. G. & Jenkins, N. A. (1991) *Trends Genet.* **7**, 113–118.
- Dracopoli, N. C., Haines, J. L., Korf, B. R., Moir, T. D., Morton, C. C., Seidman, C. E., Seidman, J. G. & Smith, D. R. (1994) *Current Protocols in Human Genetics* (Wiley Interscience, New York).
- Kozak, M. (1996) *Mamm. Genome* **7**, 563–574.
- Hannoufa, A., Negru, V., Eisner, G. & Lemieux, B. (1996) *Plant J.* **10**, 459–467.
- Borden, K. L. & Freemont, P. S. (1996) *Curr. Opin. Struct. Biol.* **6**, 395–401.
- Yu, H., Peters, J. M., King, R. W., Page, A. M., Hieter, P. & Kirschner, M. W. (1998) *Science* **279**, 1219–1222.
- Huibregts, J. M., Scheffner, M., Beaudenon, S. & Howley, P. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2563–2567.
- Taban, C. H., Hondermarck, H., Bradshaw, R. A. & Boilly, B. (1996) *Experientia* **52**, 865–870.
- Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., et al. (1995) *Cell* **81**, 27–40.
- Stewart, A. (1995) *Trends in Genetics Nomenclature Guide* (Elsevier, Cambridge, U.K.).

# The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor

Christopher Byrd, Glenn C. Turner and Alexander Varshavsky<sup>1</sup>

Division of Biology, California Institute of Technology,  
1200 East California Boulevard, Pasadena, CA 91125, USA

<sup>1</sup>Corresponding author  
e-mail: avarsh@cco.caltech.edu

Ubiquitin-dependent proteolytic systems underlie many processes, including the cell cycle, cell differentiation and responses to stress. One such system is the N-end rule pathway, which targets proteins bearing destabilizing N-terminal residues. Here we report that Ubr1p, the main recognition component of this pathway, regulates peptide import in the yeast *Saccharomyces cerevisiae* through degradation of Cup9p, a 35 kDa homeodomain protein. Cup9p was identified using a screen for mutants that bypass the previously observed requirement for Ubr1p in peptide import. We show that Cup9p is a short-lived protein ( $t_{1/2} \sim 5$  min) whose degradation requires Ubr1p. Cup9p acts as a repressor of *PTR2*, a gene encoding the transmembrane peptide transporter. In contrast to engineered N-end rule substrates, which are recognized by Ubr1p through their destabilizing N-terminal residues, Cup9p is targeted by Ubr1p through an internal degradation signal. The Ubr1p-Cup9p-Ptr2p circuit is the first example of a physiological process controlled by the N-end rule pathway. An earlier study identified Cup9p as a protein required for an aspect of resistance to copper toxicity in *S. cerevisiae*. Thus, one physiological substrate of the N-end rule pathway functions as both a repressor of peptide import and a regulator of copper homeostasis.

**Keywords:** *CUP9/N-end rule/peptide import/proteolysis/PTR2/UBR1*

## Introduction

Many regulatory proteins are short-lived *in vivo* (Schwob *et al.*, 1994; King *et al.*, 1996; Varshavsky, 1996). This metabolic instability makes possible rapid adjustment of the protein's concentration (or subunit composition) through changes in the rates of its synthesis or degradation. Protein degradation plays a role in a multitude of processes, including cell growth, division, differentiation and responses to stress. In eukaryotes, a large fraction of intracellular proteolysis is mediated by the ubiquitin system. Ubiquitin (Ub) is a 76-residue protein whose covalent conjugation to other proteins marks them for processive degradation by the 26S proteasome—an ATP-dependent, multisubunit protease (Jentsch and Schlenker, 1995; Hilt and Wolf, 1996; Hochstrasser, 1996; Rubin *et al.*, 1997).

Features of proteins that confer metabolic instability are called degradation signals (degrons). One of the degradation signals recognized by the ubiquitin system is called the N-degron. It comprises two essential determinants: a destabilizing N-terminal residue and an internal lysine of a substrate (Bachmair *et al.*, 1986; Varshavsky, 1996). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989). A set of N-degrons bearing different N-terminal residues that are destabilizing in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria (Varshavsky, 1996).

The N-end rule pathway is organized hierarchically. In eukaryotes such as *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase (Nt-amidase), into the secondary destabilizing residues Asp and Glu (Baker and Varshavsky, 1995). Secondary residues, in turn, function through their conjugation to Arg by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase) (Balzi *et al.*, 1990). Arg is one of several primary destabilizing N-terminal residues which are bound directly by N-recognin, a 225 kDa E3 protein encoded by the *UBR1* gene (Bartel *et al.*, 1990). Ubr1p, together with the associated ubiquitin-conjugating (E2) enzyme Ubc2p, mediates the formation of a substrate-linked multiubiquitin chain (Varshavsky, 1996).

The N-end rule pathway was first encountered in experiments that explored, in *S. cerevisiae*, the metabolic fate of a fusion between Ub and a reporter such as *Escherichia coli*  $\beta$ -galactosidase ( $\beta$ -gal) (Bachmair *et al.*, 1986). While such engineered N-end rule substrates have been extensively characterized (Varshavsky, 1996), little is known about their physiological counterparts. The few identified so far include the *GPA1*-encoded G $\alpha$  subunit of the *S. cerevisiae* heterotrimeric G protein, which mediates the pheromone response in this fungus, and RNA polymerases of alphaviruses whose hosts include mammalian and insect cells (de Groot *et al.*, 1991; Madura and Varshavsky, 1994). Physiological functions of the instability of these proteins remain to be understood (Varshavsky, 1996). Inactivation of the N-end rule pathway in *S. cerevisiae*—through deletion of the *UBR1* gene—results in cells which grow slightly slower than their wild-type counterparts, and are impaired in sporulation (increased frequency of ascospores containing fewer than four spores), but otherwise appear to be normal (Bartel *et al.*, 1990).

Recently, Becker and colleagues (Alagramam *et al.*, 1995) have reported that *ubr1* $\Delta$  cells are deficient in the

A	Strain	Genotype	Peptide import	Strain	Genotype	Peptide import
	JD53	wt	+	MHY599	<i>pas2Δ</i>	+
	JD83-1A	<i>ubr1Δ</i>	-	CBY8	<i>ubc2Δ, ubc1Δ</i>	+
	KM207-1	<i>ate1Δ</i>	+	CBY11	<i>ubc2Δ, cdc34-1<sup>1</sup></i>	+
	SGY6	<i>nta1Δ</i>	+	EJY105/6	<i>ubc2Δ, ubc4Δ</i>	-
	CBY20	<i>ptr2Δ</i>	+	CBY4	<i>ubc2Δ, ubc5Δ</i>	+
	EJ2	<i>ubc1Δ</i>	+	CBY5	<i>ubc2Δ, ubc6Δ</i>	+
	BBY67	<i>ubc2Δ</i>	+/-	CBY6	<i>ubc2Δ, ubc7Δ</i>	+
	JD34	<i>ubc2Δ</i>	+/-	CBY7	<i>ubc2Δ, ubc8Δ</i>	+
	EJY102	<i>ubc2Δ</i>	+/-	CBY9	<i>ubc2Δ, ubc9-1<sup>1</sup></i>	+
	RJD549	<i>cdc34-1<sup>1</sup></i>	+	CBY10	<i>ubc2Δ, pas2Δ</i>	+
	RJD795	<i>cdc34-2<sup>1</sup></i>	+	EJ10	<i>ubc4Δ, ubc5Δ</i>	+
	EJ3	<i>ubc4Δ</i>	+	MHY503	<i>ubc4Δ, ubc6Δ</i>	+
	EJ1	<i>ubc5Δ</i>	+	MHY554	<i>ubc4Δ, ubc7Δ</i>	+
	MHY495	<i>ubc6Δ</i>	+	MHY552	<i>ubc6Δ, ubc7Δ</i>	+
	MHY507	<i>ubc7Δ</i>	+	MHY550	<i>ubc4Δ, 5Δ, 6Δ</i>	+
	MHY601	<i>ubc8Δ</i>	+	MHY557	<i>ubc4Δ, 6Δ, 7Δ</i>	+
	YWO55	<i>ubc9-1<sup>1</sup></i>	+	MHY570	<i>ubc4Δ, 5Δ, 6Δ, 7Δ</i>	+

**Fig. 1.** The import of peptides is decreased in the absence of Ubc2p and virtually abolished in the absence of Ubc2p and Ubc4p. *Saccharomyces cerevisiae* mutants deficient in one or more ubiquitin-conjugating (E2) enzymes or other components of the ubiquitin system were tested for their ability to import peptides, using the halo assay and a toxic dipeptide L-leucyl-L-ethionine (Leu-Eth) (see Materials and methods). ‘+’, ‘±’ and ‘-’ denote, respectively, the apparently wild-type, significantly reduced, and undetectable levels of peptide import. A superscript ‘1’ refers to strains that carried a *ts* allele of an essential E2 enzyme, either Cdc34p (Ubc3p) or Ubc9p; the 30°C temperature of the test was semi-permissive for these strains. (A) Summary of the results. (B–F) Examples of the actual halo assays, with wild-type (B), *ptr2Δ* (C), *ubr1Δ* (D), *ubc2Δ* (E), *ubc4Δ* (F) and *ubc2Δ ubc4Δ* (G) strains of *S.cerevisiae*. Elimination of some E2 enzymes in the *ubc2Δ* background restored halo formation, presumably because such strains were growth-impaired in a way that made them hypersensitive to the toxicity of ethionine. By contrast, although *ubc2Δ ubc4Δ* cells were also growth-impaired, they were import-defective and therefore grew in the immediate vicinity of the filter.

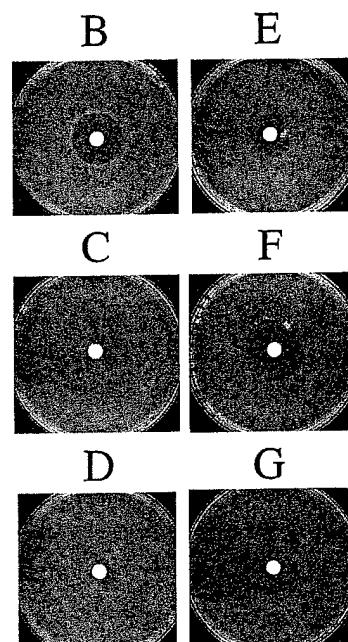
import of di- and tripeptides, suggesting that this process, which is universal among living cells, requires the N-end rule pathway. In the present work, we identified the underlying regulatory mechanism and discovered a new physiological substrate of the N-end rule pathway, the homeodomain protein Cup9p. This short-lived protein is targeted for degradation by Ubr1p, and acts as a transcriptional repressor of *PTR2*, a gene that encodes a transmembrane peptide transporter. The Ubr1p–Cup9p–Ptr2p circuit is the first example of a physiological process controlled by the N-end rule pathway.

## Results

### The involvement of Ubc2p and Ubc4p E2 enzymes in the control of peptide import

We began by asking whether components of the N-end rule pathway other than Ubr1p were also necessary for the import of peptides. Previous work has shown that Ubc2p, one of 13 ubiquitin-conjugating (E2) enzymes of *S.cerevisiae*, is required for the degradation of engineered N-end rule substrates, and is physically associated with the E3 protein Ubr1p (N-recognin) (Jentsch, 1992; Madura *et al.*, 1993).

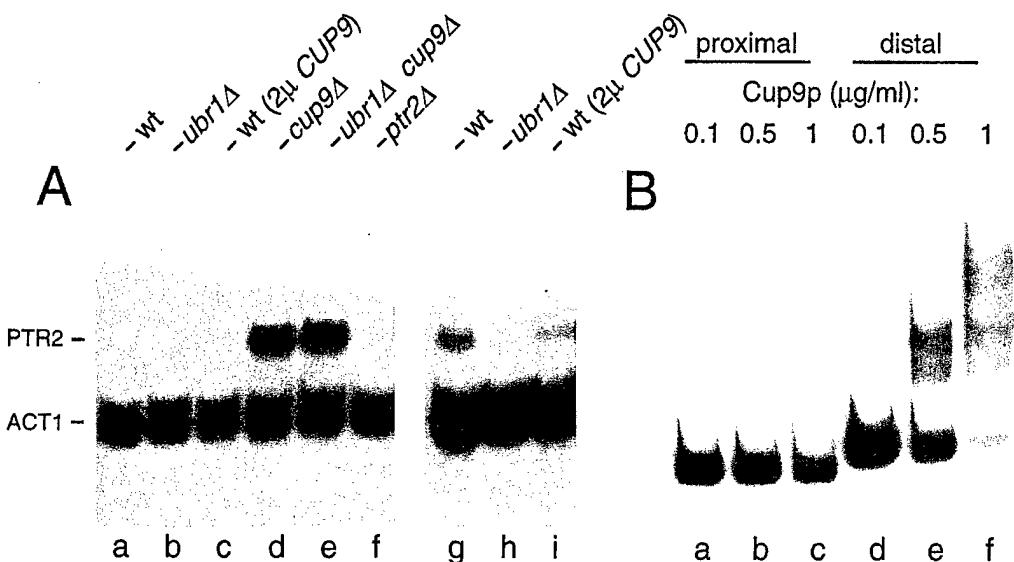
To test for the ability of *S.cerevisiae* to import peptides, we used a halo assay, in which a filter soaked in the toxic dipeptide L-leucyl-L-ethionine (Leu-Eth) is placed on a plate, inhibiting the growth of import-competent cells near the filter. By this test, the elimination of *UBC2* impaired, but did not abolish, the import of peptides (Figure 1A, B and E). To determine which of the other E2 enzymes, if



any, were required for the residual peptide import observed in *ubc2Δ* cells, a number of single and multiple mutants in *UBC* genes were examined (Figure 1). We found that the elimination of both *UBC2* and *UBC4* virtually abolished the import of peptides (Figure 1A, B and E–G). Elimination of Ubc4p, one of the more abundant E2 enzymes (Bachmair *et al.*, 1986; Jentsch, 1992), had previously been noticed to decrease slightly the activity of the N-end rule pathway (Bartel, 1990). Cells lacking either Nta1p or Atel1p—the ‘upstream’ components of this pathway—were also examined and found unimpaired in the import of peptides, in contrast to *ubr1Δ* and *ubc2Δ ubc4Δ* cells (Figure 1A).

### Identification of Cup9p as a negative regulator of peptide import

Previous work (Alagramam *et al.*, 1995) has shown that deletion of *UBR1* greatly reduces the level of *PTR2* mRNA, which encodes the transmembrane peptide transporter. This result, and our observation that peptide import requires the presence of at least one of two specific ubiquitin-conjugating enzymes (Figure 1), suggested a model in which expression of the Ptr2p transporter is regulated by a short-lived repressor that is degraded by the N-end rule pathway. In cells lacking *UBR1*, the repressor would be expected to accumulate, thereby blocking peptide import. One prediction of this model is that inactivation of this repressor would bypass the requirement for Ubr1p in peptide import. A screen for such ‘bypass’ mutations (see Materials and methods) yielded 199 recessive isolates, of which 101 defined one complementation group, termed



**Fig. 2.** Cup9p is a repressor of the *PTR2* gene. (A) Expression of the *S.cerevisiae* *PTR2* gene in different genetic backgrounds. Equal amounts of total RNA isolated from different strains were analyzed by Northern hybridization (see Materials and methods), using the *PTR2* (peptide transporter) and *ACT1* (actin) genes as  $^{32}$ P-labeled probes. Lane a, JD52 (*CUP9 UBR1*) (wild-type) transformed with pCB201 (empty vector). Lane b, JD55 (*CUP9 ubr1Δ*) transformed with pCB201. Lane c, JD52 (*CUP9 UBR1*) transformed with the high-copy plasmid pCB209 that expressed *CUP9* from its natural promoter. Lane d, CBY18 (*cup9Δ UBR1*). Lane e, CBY16 (*cup9Δ ubr1Δ*). Lane f, CBY21 (*CUP9 UBR1 ptr2Δ*). Lanes g–i, same as lanes a–c, but a longer autoradiographic exposure to highlight the tight repression of *PTR2* by Cup9p in *ubr1Δ* cells (b, h) and the difference between levels of *PTR2* mRNA in the wild-type cells (a, g) and their counterparts that overexpressed Cup9p (c, i). (B) Cup9p specifically binds to a site in the  $P_{PTR2}$  promoter. A gel shift assay with Cup9p-H<sub>6</sub>, poly-dI-dC and  $^{32}$ P-labeled DNA fragments of the *PTR2* promoter hybridization (see Materials and methods). Lanes a–c, with a fragment (-1 to -447) proximal to the inferred start codon of the *PTR2* ORF. Lanes d–f, same as lanes a–c, but with a more distal DNA fragment (-448 to -897). Concentrations of Cup9p-H<sub>6</sub> (in  $\mu$ g/ml) are indicated above the lanes.

*sub1* (suppressor of a block to peptide import in *ubr1Δ*). In agreement with the model's prediction, *sub1* mutants acquired the ability to express *PTR2* in the absence of *UBR1* (Figure 2A, lane b versus lane e). The *sub1* locus was cloned by complementation (see Materials and methods), and was found to be the *CUP9* gene.

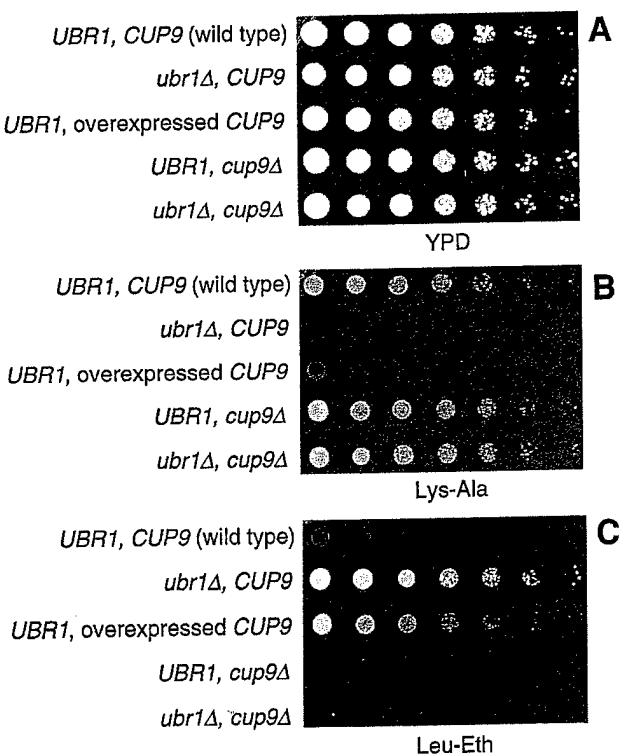
*CUP9* was originally identified by Knight *et al.* (1994) as a gene whose disruption impairs the copper resistance of *S.cerevisiae* growing on lactate, a non-fermentable carbon source. Under these conditions, Cup9p plays a major (but mechanistically obscure) role in copper homeostasis (Knight *et al.*, 1994). *CUP9* encodes a 35 kDa protein that contains a homeodomain, an ~60-residue helix-turn-helix DNA-binding motif present in many eukaryotic regulatory proteins (Wolberger, 1996). Outside the homeodomain region, the sequence of Cup9p is not similar to sequences in databases.

To verify that *CUP9* and *SUB1* were the same gene, complementation tests were carried out. Two independently derived *ubr1Δ cup9::LEU2* strains (CBY16 and CBY17) were crossed to *ubr1Δ sub1-1* (CBY15), and the resulting diploids (CBY23 and CBY24, respectively) were tested for their ability to import dipeptides (see Materials and methods). As would be expected of allelic loci, *cup9::LEU2* and *sub1-1* failed to complement one another: both diploid strains remained import-competent (data not shown). In another test, CBY23 and CBY24 were sporulated, and the segregants were analyzed for the presence of the *LEU2* gene (integrated at the *CUP9* locus) and for the ability to import peptides. Among the eight tetrads tested, *LEU2* was present in two of the four segregants, whereas all four segregants were invariably import-competent, a pattern expected if *CUP9* and *SUB1* were one and the same gene.

To examine the regulation of peptide import by Cup9p and Ubr1p, congenic *S.cerevisiae* strains that lacked, expressed or overexpressed Cup9p and/or Ubr1p were constructed and assayed for peptide import by growth on selective media. Suspensions of cells (auxotrophic for lysine) were serially diluted and plated on either rich media, minimal media lacking lysine and containing Lys-Ala dipeptide (selecting for peptide import), or minimal media containing the toxic dipeptide Leu-Eth (selecting against peptide import). All strains grew at comparable rates on rich media (Figure 3A). On minimal media that supplied the essential lysine as the Lys-Ala dipeptide, the *CUP9 UBR1* (wild-type), *cup9Δ UBR1* and *cup9Δ ubr1Δ* strains grew at comparable rates (Figure 3B), whereas the *CUP9 ubr1Δ* strain failed to grow (Figure 3B), in agreement with the observation that *UBR1* is required for peptide import (Figure 1). Opposite growth patterns were observed on media containing toxic dipeptide (Figure 3C). Comparison of the data in Figure 3 with *PTR2* mRNA levels (Figure 2A) suggested that *CUP9* exerts its effect on the import of peptides by repressing transcription of *PTR2*. For example, wild-type (*CUP9 UBR1*) cells overexpressing Cup9p exhibited reduced levels of *PTR2* mRNA (Figure 2A) and decreased sensitivity to toxic dipeptide (Figure 3C), whereas strains lacking *CUP9* (*cup9Δ UBR1* and *cup9Δ ubr1Δ*) overexpressed *PTR2* (Figure 2A) and were hypersensitive to toxic dipeptide (Figure 3C).

#### ***Cup9p is a repressor of the *PTR2* gene***

To address the mechanism of repression by Cup9p, we asked whether purified Cup9p (see Materials and methods) could selectively bind to specific regions of the *PTR2* promoter. Gel shift assays in the presence of poly-dI-dC competitor DNA were performed with labeled DNA



**Fig. 3.** Relative capacity for peptide import in congeneric *S.cerevisiae* strains that contained, lacked or overexpressed Cup9p and/or Ubr1p. Serial dilutions of the indicated strains were deposited by a 48-pin applicator onto either rich (YPD) medium (A), minimal medium containing 66  $\mu$ M Lys-Ala dipeptide as the sole source of lysine (B), or minimal medium containing both lysine (at 110  $\mu$ M) and the toxic dipeptide Leu-Eth (at 55  $\mu$ M) (C). The plates were incubated at 30°C for 1–2 days. The relevant genetic loci are shown on the left. The 'UBR1, overexpressed CUP9' strain (JD52-2 $\mu$ -CUP9) carried the high-copy plasmid pCB209 that expressed Cup9p from its natural promoter.

fragments of the *PTR2* promoter and His<sub>6</sub>-tagged Cup9p (Cup9p-H<sub>6</sub>) purified from *E.coli*. Cup9p-H<sub>6</sub> bound to a site within a distal region of the *PTR2* promoter (positions -448 to -897 relative to the inferred *CUP9* start codon), but did not bind to the proximal region of the *PTR2* promoter (positions -1 to -447) under the same conditions (Figure 2B).

The transcriptional repressor function of Cup9p was further suggested by the finding that the co-repressor complex Tup1p/Ssn6p plays a role in the control of peptide import. The Tup1p/Ssn6p complex inhibits transcription of many yeast genes through interactions with gene-specific DNA-binding repressors such as Mat02p (Chen *et al.*, 1993; Smith *et al.*, 1995), a homeodomain homolog of Cup9p. We found that most of our *sub* mutants that were not *CUP9* mutants could be complemented by low-copy plasmids bearing *SSN6* or *TUP1* (G.Turner, S.Saha and A.Varshavsky, unpublished data). In addition, deletion of *SSN6*, like deletion of *CUP9*, restored the ability of *ubr1Δ* cells to import peptides (data not shown).

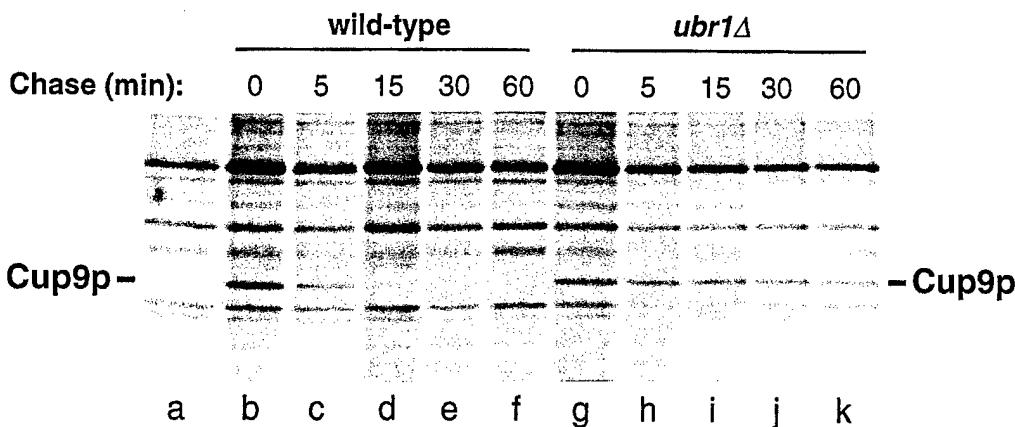
#### **Ubr1p-dependent degradation of Cup9p**

The fact that deletion of *CUP9* renders *PTR2* transcription independent of Ubr1p (Figure 2A), and the observation that overexpression of Ubr1p enhances the import of peptides in Cup9p-expressing strains (data not shown) suggested that the N-end rule pathway regulates peptide

import by targeting Cup9p for degradation. To test this conjecture, we carried out pulse-chase experiments with a C-terminally FLAG-tagged Cup9p in *UBR1* and *ubr1Δ* cells. Cup9p-FLAG was a very short-lived protein ( $t_{1/2}$  ~5 min) in *UBR1* cells (Figure 4). By contrast, Cup9p was much longer-lived ( $t_{1/2}$  >30 min) in *ubr1Δ* cells (Figure 4). Degradation of Cup9p was also found to depend upon *UBC2* and *UBC4* (data not shown), in agreement with the observation that a *ubc2Δ ubc4Δ* double mutant failed to import dipeptides (Figure 1G). The residual instability of Cup9p in *ubr1Δ* cells (Figure 4) suggested the presence of a second, Ubr1p-independent degron; this pattern is reminiscent of another homeo-domain repressor, Mat02p, which also contains at least two distinct degradation signals (Hochstrasser and Varshavsky, 1990; Chen *et al.*, 1993).

Ubr1p recognizes engineered N-end rule substrates through their destabilizing N-terminal residues (Varshavsky, 1996). To determine the N-terminal residue of Cup9p, we overexpressed and purified Cup9p-FLAG from *ubr1Δ S.cerevisiae* and subjected it to N-terminal sequencing. Cup9p-FLAG was found to have a blocked (presumably acetylated) N-terminus (see Materials and methods), suggesting that Ubr1p targets Cup9p through a degron distinct from the N-degron. Independent evidence for this conjecture was produced through the analysis of GST-Cup9p-ha<sub>2</sub>, a fusion of glutathione transferase (GST) and C-terminally ha-tagged Cup9p. Pulse-chase analysis of GST-Cup9p-ha<sub>2</sub> revealed that the fusion protein was nearly as short-lived *in vivo* as the N-terminally unmodified Cup9p-FLAG (Figure 4, lanes b–e; compare with Figure 5, lanes b–e). If Cup9p were targeted for processive degradation through a destabilizing N-terminal residue, a preliminary proteolytic cleavage(s) of Cup9p would be required to expose such a residue (Varshavsky, 1996). In the case of GST-Cup9p-ha<sub>2</sub>, this cleavage would generate a proteolytic fragment consisting of GST and an N-terminal portion of Cup9p preceding the cleavage site. Since free GST has been found to be long-lived when expressed in yeast (data not shown), accumulation of such a proteolytic fragment would be expected to accompany the degradation of GST-Cup9p-ha<sub>2</sub>.

Pulse-chase analysis of GST-Cup9p-ha<sub>2</sub>, using glutathione-agarose beads to isolate GST-containing proteins (see Materials and methods), indicated that the degradation of GST-Cup9p-ha<sub>2</sub> by the N-end rule pathway was not accompanied by the appearance of a fragment containing the N-terminal GST moiety (Figure 5). This finding strongly suggested that Cup9p bears an 'internal' degron recognized by Ubr1p. Additional support for this conjecture was provided by truncation analysis of a Cup9p-DHFR-myc fusion protein. These experiments indicated that the N-terminal 81 residues of the 306-residue Cup9p are dispensable for its Ubr1p-dependent degradation, strongly suggesting that the relevant degradation signal resides in the C-terminal two-thirds of Cup9p (C.Byrd, I.Davydov and A.Varshavsky, unpublished data). Although these results are fully consistent with the presence of an internal Ubr1p-dependent degron in Cup9p, there remains the less parsimonious possibility that Cup9p is degraded via *trans*-targeting (Johnson *et al.*, 1990). In this process, the two determinants of the N-degron (a destabilizing N-terminal residue and a ubiquitin-accepting internal Lys



**Fig. 4.** Degradation of Cup9p by the N-end rule pathway. *Saccharomyces cerevisiae* strains carrying pCB210, a low-copy plasmid that expressed Cup9p-FLAG from the *P<sub>CUP9</sub>* promoter, were labeled at 30°C with [<sup>35</sup>S]methionine/cysteine for 5 min, followed by a chase for 0, 5, 15, 30 and 60 min, preparation of extracts, immunoprecipitation of Cup9p-FLAG with a monoclonal anti-FLAG antibody, SDS-PAGE and autoradiography/quantitation (see Materials and methods). Lane a, JD55 (*ubr1Δ*) cells that carried pCB200 (vector alone). Lanes b-f, JD52 (*UBR1*) (wild-type) cells that carried pCB210, expressing Cup9p-FLAG. Lanes g-k, same as lanes b-f, but with JD55 (*ubr1Δ*) cells.

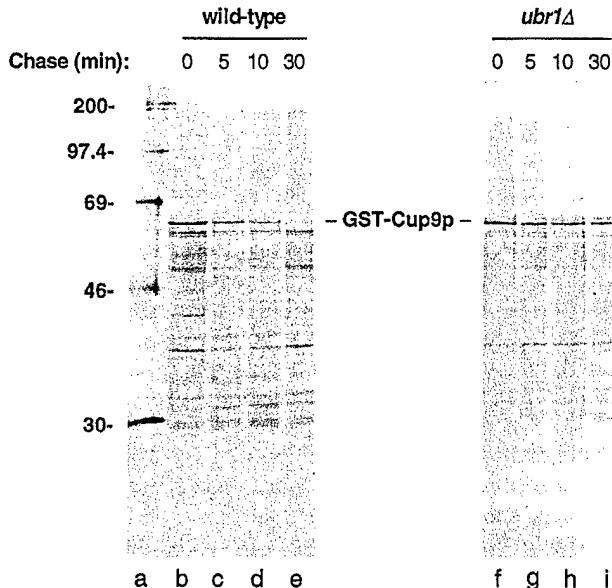
residue) reside in two different polypeptides, whose interaction yields an active N-degron, leading to ubiquitylation and degradation of the polypeptide bearing the Lys residue (Johnson *et al.*, 1990). At present, little is known about the protein ligands of Cup9p, save for the likely possibility that Cup9p interacts with the Tup1p/Ssn6p repressor complex (see above), similarly to the previously established interaction of this complex with Matα2, a homeodomain-containing homolog of Cup9p (Smith *et al.*, 1995).

## Discussion

The discovery that Ubr1p controls the import of peptides through degradation of the Cup9p repressor (Figure 6) identifies the first clear physiological function of the N-end rule pathway. The existence of mammalian, plant and bacterial homologs of the yeast Ptr2p transporter suggests that the import of peptides in these organisms may also be regulated by the N-end rule pathway.

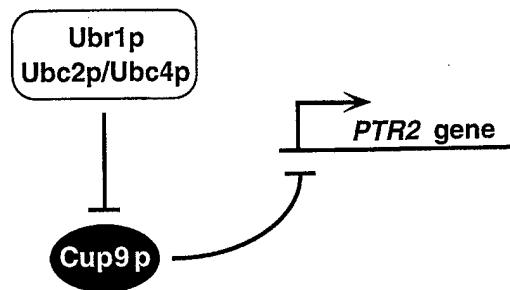
Why was this pathway, rather than another Ub-dependent proteolytic system, recruited in the course of evolution for the regulation of peptide import? A plausible explanation is suggested by the ability of Ubr1p to bind peptides bearing destabilizing N-terminal residues (Varshavsky, 1996). Since more than half of the 20 amino acids are destabilizing in the yeast and mammalian N-end rules (Varshavsky, 1996), a significant fraction of imported peptides would be expected to compete with Cup9p for binding to Ubr1p. This competition may decrease the rate of Cup9p degradation. [Dipeptides added to a culture of *S.cerevisiae* have been shown to inhibit the N-end rule pathway (Baker and Varshavsky, 1991).] The ensuing increase in the level of Cup9p repressor would in turn decrease the production of the Ptr2p transporter, diminishing the rate of peptide import. This 'peptide-sensing' negative feedback loop would maintain the intracellular concentration of short peptides within a predetermined range—potentially a useful feature, made possible by the substrate-binding properties of Ubr1p. Experiments to verify this model are under way.

Another interesting aspect of the Ubr1p-Cup9p-Ptr2p



**Fig. 5.** Degradation of GST-Cup9p-ha<sub>2</sub> is not preceded by a processing that releases a GST-containing fragment. *Saccharomyces cerevisiae* strains carrying pCB210, a low-copy plasmid that expressed GST-Cup9p-ha<sub>2</sub> from the *P<sub>GAL1</sub>* promoter, were labeled at 30°C with [<sup>35</sup>S]methionine/cysteine for 5 min, followed by a chase for 0, 5, 10 and 30 min, preparation of extracts, isolation of GST-Cup9p-ha<sub>2</sub> on glutathione-agarose beads, SDS-PAGE and autoradiography/quantitation (see Materials and methods). Lanes b-e, JD52 (*UBR1*) (wild-type) cells expressing GST-Cup9p-ha<sub>2</sub>. Lanes f-i, same as lanes b-e, but with JD55 (*ubr1Δ*) cells. Lane a, molecular mass markers (in kDa).

regulatory circuit is its possible relevance to copper homeostasis. In addition to functioning as a repressor of peptide transport, Cup9p also contributes to the resistance of *S.cerevisiae* to copper toxicity during growth on lactate, a non-fermentable carbon source (Knight *et al.*, 1994). Copper homeostasis is mediated by a complex set of circuits, some of which also regulate iron metabolism (Zhou and Thiele, 1993; Ooi *et al.*, 1996). The double role of Cup9p in regulating both peptide import and copper homeostasis may signify a physiological connection between these seemingly unrelated processes.



**Fig. 6.** A model for regulation of peptide import in *S.cerevisiae*. The expression of *PTR2*, which encodes a transmembrane peptide transporter, is regulated by the short-lived, homeodomain-containing transcriptional repressor Cup9p. The concentration of Cup9p is controlled in part through its degradation by the N-end rule pathway, whose targeting components include Ubr1p and either Ubc2p or Ubc4p (see the main text).

Our work adds Cup9p to the list of short-lived regulatory proteins whose degradation is mediated by the ubiquitin system. Many of these proteins are negative regulators. For example, Mat $\alpha$ 2p, a homeodomain-containing homolog of Cup9p, controls the mating type of *S.cerevisiae* ( $\alpha$  or  $\alpha'$ ) through repression of  $\alpha$ -specific genes (Herskowitz, 1989). Mat $\alpha$ 2p appears to be constitutively short-lived in haploid cells (Hochstrasser and Varshavsky, 1990). Therefore, cessation of Mat $\alpha$ 2p synthesis upon the conversion of an  $\alpha$  cell into an  $\alpha'$  cell results in rapid disappearance of Mat $\alpha$ 2p and the establishment of  $\alpha$ -specific circuits (Hochstrasser, 1996). Progression of the cell cycle is also controlled by short-lived negative regulators, in particular by Sic1p, an inhibitor of CDK, the cyclin-dependent kinase (Schwob *et al.*, 1994). In this case, however, a rapid 'on-switch' is based on the phosphorylation-induced degradation of the previously stable Sic1p (King *et al.*, 1996). Whether the Cup9p repressor (Figure 6) is constitutively short-lived or whether its stability is regulated by external conditions such as, for example, different nitrogen sources, remains to be determined.

The finding that Cup9p lacks a destabilizing N-terminal residue indicates that Ubr1p, the recognition component of the N-end rule pathway, is able to target substrates bearing either internal degradation signals or N-degrons. The G $\alpha$  subunit of the G protein—the other known physiological substrate of Ubr1p in *S.cerevisiae*—also lacks a destabilizing N-terminal residue (Madura and Varshavsky, 1994). Thus, a substantial fraction of naturally short-lived proteins targeted by the N-end rule pathway may bear internal degrons rather than N-degrons, a possibility that the identification of other physiological substrates of this pathway will address.

## Materials and methods

### Strains, media and genetic techniques

*Saccharomyces cerevisiae* strains were grown in rich (YPD) medium containing 2% peptone, 1% yeast extract and 2% glucose or in synthetic yeast media, containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients at concentrations specified by Sherman *et al.* (1986) and either 2% glucose (SD medium), 2% raffinose (SM-raffinose medium) or 2% galactose (SM-galactose medium) as carbon sources. SHM-glucose medium used in halo assays was prepared according to Island *et al.* (1987) and was identical to SD medium except that it lacked methionine and contained allantoin (1 mg/ml) and yeast nitrogen base (Difco, 1.7 mg/ml) without amino acids and ammonium

sulfate. *Escherichia coli* strain DH5 $\alpha$  was used for plasmid propagation and cloning steps. For induction of the *P<sub>GAL1</sub>* promoter, cells were grown to an *A<sub>600</sub>* of 0.5–1 in SM-raffinose medium, pelleted and transferred to SM-galactose medium for 3 h.

### Halo and dilution assays

Peptide import was assayed using the halo and dilution methods. For halo assays (Island *et al.*, 1987), cells were grown to an *A<sub>600</sub>* of ~1 in SHM-glucose medium with auxotrophic supplements. Cells were pelleted by centrifugation and resuspended in water to ~5 × 10<sup>7</sup> cells/ml. Then 0.1 ml of cell suspension was mixed with 10 ml of 0.8% noble agar (Difco) at 55°C, and spread on plates containing 20 ml of SHM medium. Sterile filter disks containing the toxic dipeptide L-leucyl-L-ethionine (Leu-Eth; 5 µmol) were placed in the middle of each plate, followed by incubation at 30°C for 1–2 days.

For dilution assays, strains were grown (under selection for plasmids) in SD medium to an *A<sub>600</sub>* of ~1. Cells from each sample (1.5 × 10<sup>7</sup>) were spun down and resuspended in 1 ml of water. The samples were serially diluted by 4-fold in a microtiter plate (150 µl/well) to generate eight different initial concentrations of cells that ranged from 1.5 × 10<sup>7</sup> to 915 cells/ml. The suspensions were spotted to various media, using a 48-pin applicator. The plates were incubated at 30°C for 1–2 days.

Leu-Eth was synthesized using standard methods of organic chemistry. Briefly, L-ethionine methyl ester (Eth-OMe) was produced from L-ethionine and methanol. Eth-OMe was then coupled with *N*-tert-butoxy-carbonyl-L-leucine-*p*-nitrophenyl ester (*N*-t-BOC-L-leucine-PNP), yielding *N*-t-BOC-L-leucyl-L-ethionine methyl ester, which was purified by flash chromatography on a silica column. *N*-t-BOC-L-leucyl-L-ethionine methyl ester was then converted into *N*-t-BOC-L-leucyl-L-ethionine by treatment with KOH. *N*-t-BOC-L-leucyl-L-ethionine was deprotected with trifluoroacetic acid, yielding Leu-Eth.

### A screen for import-competent mutants in the *ubr1Δ* background

Thirty 5-ml cultures of JD83-1A (*leu2-3 ubr1Δ*), auxotrophic for leucine, were grown to an *A<sub>600</sub>* of ~1, and ~1.5 × 10<sup>7</sup> cells from each culture were plated onto SD medium lacking leucine and histidine but containing the leucyl-histidine dipeptide at 0.23 µM. Plates were incubated at 30°C for 2 days, selecting for mutants able to grow on this medium. Of the 320 sub mutants ('suppressors of a block to peptide import in *ubr1Δ*') thus obtained, 199 were found by complementation tests to be clearly recessive. Among these, 101 mutants belonged to one complementation group, termed *sub1*.

### Isolation of the *CUP9* (*SUB1*) gene

A 50 ml culture of *S.cerevisiae* CBY15 (*ubr1Δ sub1-1*) was grown to an *A<sub>600</sub>* of ~1 in SHM-glucose. Cells were made competent with lithium acetate (Ausubel *et al.*, 1992), and 25 0.1-ml samples (~7 × 10<sup>7</sup> cells/ml) were transformed with a yeast genomic DNA library (American Type Culture Collection; #77164) in the *TRP1*, *CEN6*-based vector pRS200 (Sikorski and Hieter, 1989). The cells were pelleted, and each sample was resuspended in 1.5 ml of SHM-glucose and incubated at 30°C for 3 h. The cells were pelleted again, each sample was resuspended in 0.1 ml of water, added to 10 ml of 0.8% Noble agar at 55°C, and spread onto SHM-glucose plates lacking Trp and containing Leu-Eth at 37 µM. Of the ~2.5 × 10<sup>4</sup> Trp<sup>+</sup> transformants plated, 14 could grow in the presence of Leu-Eth. Of these, two yielded the library-derived plasmids, pSUB1-1 and pSUB1-6, that complemented the *sub1* mutation of CBY15. The insert of pSUB1-6 contained the insert of pSUB1-1. Partial sequencing of the ~8.2 kb insert of pSUB1-1 identified it as a region of the *S.cerevisiae* chromosome XVI. Deletion analysis (not shown) localized the complementing activity to an ~2.8 kb *HindIII-KpnI* fragment, whose sequence revealed the presence of two ORFs (*CUP9* and *SCYPL178W*). Further deletion analysis (not shown) localized the complementing activity to the previously isolated (Knight *et al.*, 1994) *CUP9* gene. Verification that *CUP9* and *SUB1* were one and the same gene was carried out as described in Results.

### Construction of null *cup9* mutants

The ~2.8 kb *HindIII-KpnI* genomic DNA fragment containing *CUP9* was ligated to *HindIII-KpnI*-cut pRS306ΔSpeI, yielding pCB117. pRS306ΔSpeI was derived from the *URA3*-bearing pRS306 (Sikorski and Hieter, 1989) through elimination of its *SpeI* site. The *CUP9* ORF was disrupted by inserting an ~2 kb, blunt-ended, *LEU2*-containing *SalI* fragment from pJJ283 into the *SpeI* site of *CUP9*, yielding pCB119, in which the *LEU2* and *CUP9* ORFs were oriented in opposite directions. An ~5 kb *HindIII-PvuII* fragment of pCB119 containing the *cup9::LEU2*

**Table I.** *Saccharomyces cerevisiae* strains used in this study

Strain	Genotype	References
DF5	MAT $\alpha$ /MAT $\alpha$ trp1-1/trp1-1 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801 gal/gal	Finley <i>et al.</i> (1987)
EJ1	MAT $\alpha$ ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
EJ2	MAT $\alpha$ ubc1Δ::URA3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
EJ3	MAT $\alpha$ ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
EJ10	MAT $\alpha$ ubc4Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
EJY102	MAT $\alpha$ ubc2Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
EJY105	MAT $\alpha$ ubc2Δ::LEU2 ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
EJY106	MAT $\alpha$ ubc2Δ::LEU2 ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of DF5 <sup>a</sup>
MHY495	MAT $\alpha$ ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY503	MAT $\alpha$ ubc4-Δ1::HIS3 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY507	MAT $\alpha$ ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY550	MAT $\alpha$ ubc4-Δ2::TRP1 ubc5-Δ1::LEU2 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY552	MAT $\alpha$ ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY554	MAT $\alpha$ ubc4-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY557	MAT $\alpha$ ubc4-Δ1::HIS3 ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY570	MAT $\alpha$ ubc4-Δ2::TRP1 ubc5-Δ1::LEU2 ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>b</sup>
MHY599	MAT $\alpha$ pas2 trp ura3-52 ade1 leu2-3	Chen <i>et al.</i> (1993)
MHY601	MAT $\alpha$ ubc8::URA3 trp1-1 ura3-1 ade2-1 his3-11 leu2-3,112 can1-100	Chen <i>et al.</i> (1993)
YWO55	MAT $\alpha$ ubc9-1 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5 <sup>c</sup>
YPH500	MAT $\alpha$ trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801	Sikorski and Hieter (1989)
BBY67	MAT $\alpha$ ubc2Δ::LEU2 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801	Derivative of YPH500 <sup>d</sup>
KM207-1	MAT $\alpha$ atelΔ::TRP1 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801	Derivative of YPH500 <sup>e</sup>
JD34	MAT $\alpha$ ubc2Δ::URA3 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801	Derivative of YPH500 <sup>f</sup>
JD51	MAT $\alpha$ /MAT $\alpha$ trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801	Dohmen <i>et al.</i> (1995)
JD52	MAT $\alpha$ trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Johnson <i>et al.</i> (1995)
JD53	MAT $\alpha$ trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Dohmen <i>et al.</i> (1995)
JD55	MAT $\alpha$ ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Madura and Varshavsky (1994)
JD83-1A	MAT $\alpha$ ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of JD51 <sup>f</sup>
RJD549	MAT $\alpha$ cdc34-1 trp1 ura3-52 leu2-3	R.Deshaires <sup>g</sup>
RJD795	MAT $\alpha$ cdc34-2 trp1 ura3-52 leu2-3	R.Deshaires <sup>g</sup>
SGY6	MAT $\alpha$ nta1Δ::TRP1 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-3,112 lys2-801	S.Groryev <sup>h</sup>
CBY4	MAT $\alpha$ ubc2Δ::URA3 ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of EJ1
CBY5	MAT $\alpha$ ubc2Δ::LEU2 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of MHY495
CBY6	MAT $\alpha$ ubc2Δ::URA3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of MHY507
CBY7	MAT $\alpha$ ubc2Δ::LEU2 ubc8::URA3 trp1-1 ura3-1 ade2-1 his3-11 leu2-3,112 can1-100	Derivative of MHY601
CBY8	MAT $\alpha$ ubc2Δ::LEU2 ubc1Δ::URA3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal	Derivative of EJ2
CBY9	MAT $\alpha$ ubc2Δ::LEU2 ubc9-1 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of YWO55
CBY10	MAT $\alpha$ ubc2Δ::LEU2 pas2 trp ura3-52 ade1 leu2-3	Derivative of MHY599
CBY11	MAT $\alpha$ ubc2Δ::LEU2 cdc34-1 trp1 ura3-52 leu2-3	Derivative of RJD549
CBY15	MAT $\alpha$ sub1-1 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of JD83-1A
CBY16	MAT $\alpha$ cup9::LEU2 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of JD55
CBY17	MAT $\alpha$ cup9::LEU2 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of JD55
CBY19	MAT $\alpha$ cup9::LEU2 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of JD52
CBY23	MAT $\alpha$ /MAT $\alpha$ sub1-1/cup9Δ::LEU2 ubr1Δ::HIS3 trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801	Produced by mating CBY15 and CBY16
CBY24	MAT $\alpha$ /MAT $\alpha$ sub1-1/cup9Δ::LEU2 ubr1Δ::HIS3 trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801	Produced by mating CBY15 and CBY17

<sup>a</sup>Johnson *et al.* (1992, 1995). A gift from E.Johnson, the Rockefeller University, New York, NY 10021-6399, USA.<sup>b</sup>Chen *et al.* (1993). A gift from M.Hochstrasser, Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA.<sup>c</sup>A gift from S.Jentsch, ZMBH, Universität Heidelberg, 69120 Heidelberg, Germany.<sup>d</sup>Dohmen *et al.* (1990).<sup>e</sup>A gift from K.Madura, Department of Biochemistry, UMDNJ-Johnson Medical School, Piscataway, NJ 08854, USA.<sup>f</sup>A gift from J.Dohmen, Heinrich-Heine-Universität, Institut für Mikrobiologie, 40225 Düsseldorf, Germany.<sup>g</sup>A gift from R.Deshaires, Division of Biology, Caltech, Pasadena, CA 91125, USA.<sup>h</sup>A gift from S.Groryev, Department of Biology, University of Massachusetts, Amherst, MA 01003, USA.

disruption allele was used to replace the wild-type *CUP9* alleles of JD52 (wild-type) and JD55 (*ubr1Δ*) by homologous recombination (Rothstein, 1991), generating strains CBY19 and CBY17, respectively.

#### CUP9-expressing plasmids

The plasmid pCB116, which expressed Cup9p-FLAG from the *P<sub>GAL1</sub>* promoter, was constructed by subcloning an ~1 kb *Bam*HI-*Eco*RI fragment containing the *CUP9*-FLAG ORF into the *Bam*HI-*Eco*RI site(s) of p416GAL1 (Mumberg *et al.*, 1994). The *CUP9*-FLAG-containing fragment was constructed by PCR amplification of the *CUP9* ORF of

plasmid pCB111 using primers PCB1 (5'-CGCGGATCCGAATAGT-TACATTCGAAGATG-3') and PCB6 (5'-CCGGAATTCTCATTTA-TCATCATCGTCTTGTAAATCATATCAGGGTTGGATAG-3'), resulting in the addition of the 8-residue FLAG epitope, DYKDDDDK, to the C-terminus of Cup9p. pCB111 was constructed by subcloning the ~2.8 kb *Hind*III-*Kpn*I fragment of the *CUP9*-containing pSUB1-1 (see above) into *Hind*III-*Kpn*I-cut pRS316 (Sikorski and Hieter, 1989). pCB202 was constructed by subcloning an ~1 kb fragment containing the *P<sub>CUP9</sub>* promoter into *Hind*III-*Bam*HI-cut pCB201. [The ~1 kb fragment was produced by PCR from pCB111 using primers PCB8

(5'-GTGTTAGTAAGCTTGAAAGGAATGCACGTATT-3') and PCB9 (5'-CCCGCGGATCCGCATGCAACTATTCTCGAAGGGTGT-3').] pCB200 (*ARS-CEN*, *LEU2*) and pCB201 (2 $\mu$ , *LEU2*) were constructed by replacing the 517 bp *Scal*-*EcoRI* fragment of pBR322 (Ausubel *et al.*, 1992) with, respectively, the 3822 bp *Scal*-*NaeI* fragment of pRS415 (Sikorski and Hieter, 1989) and the 4650 bp *Scal*-*NaeI* fragment of pRS425 (Christianson *et al.*, 1992).

The plasmid pCB209 (2 $\mu$ , *LEU2*), which expressed *CUP9* from the *P<sub>CUP9</sub>* promoter, was constructed by replacing the *SphI*-*SaiI* fragment of pCB202 with an ~1 kb fragment containing the *CUP9* ORF that was produced by PCR from pCB111, using primers PCB10 (5'-CCC-GCGGATCCGCATGCGAAGATGAATTATACTGC-3') and PCB12 (5'-CCCGCGCGGTGACCTCAATTATCATATCAGGGTTGGAGATG-3'). pCB210 (*ARS-CEN*, *LEU2*) that expressed Cup9p-FLAG from the *P<sub>CUP9</sub>* promoter was constructed by replacing the *SphI*-*SaiI* fragment of pCB202 with an ~1 kb fragment containing the *CUP9-FLAG* ORF, which was produced from pCB111 using primers PCB10 (see above) and PCB13 (5'-CCCGCGCGGTGACCTCAATTATCATATCGTCTTGTATCATTCATATCAGGGTTGGAGATG-3'), yielding pCB211. The ~2 kb *HindIII*-*SaiI* fragment of pCB211 containing *P<sub>CUP9</sub>* and the *CUP9-FLAG* ORF was subcloned into pCB200, yielding pCB210.

Plasmid pCB120, expressing GST-Cup9p-ha<sub>2</sub> from the *P<sub>GAL1</sub>* promoter, was constructed by subcloning the ~1.6 kb *XbaI*-*EcoRI* fragment, containing the *GST-CUP9-ha<sub>2</sub>* ORF, into the *XbaI*-*EcoRI* site(s) of p416GAL1. The *XbaI*-*EcoRI* fragment was produced by PCR amplification of the *GST-CUP9-ha<sub>2</sub>* ORF of pGEX-2T-CUP9-ha<sub>2</sub>, using the primers PCB3 (5'-CCGGAATTCTCAAGCTAATCTGAAACATC-GTATGGTAAGCGTAATCTGAAACATCGTATGGTAATTCTATCAGGGTTGGATAG-3') and PCB5 (5'-TGCTCTAGAACAGT-ATTATCATGTCCCCCTATA-3'). pGEX-2T-CUP9-ha<sub>2</sub> was constructed by subcloning an ~1 kb fragment containing the *CUP9-ha<sub>2</sub>* ORF into the *BamHI*-*EcoRI* site(s) of pGEX-2T (Pharmacia), resulting in an in-frame fusion of the sequence encoding 26 kDa glutathione *S*-transferase (GST) domain of *Saccharomyces japonicum* (Smith and Johnson, 1988) to the second codon of *CUP9*. The *CUP9-ha<sub>2</sub>*-containing fragment was produced by PCR amplification of the *CUP9* ORF of pCB111 using the primers PCB3 (see above) and PCB4 (5'-CGCGGATCCAATTATACTGCGAAATACAAAAC-3'). This step added to the C-terminus of Cup9p a sequence encoding a tandem repeat of the 9-residue sequence YPYDVPDYA, derived from hemagglutinin (ha) of influenza virus.

#### **Northern hybridization**

RNA was isolated from *S.cerevisiae* as described (Schmitt *et al.*, 1990). Electrophoresis of the RNA samples was carried out on a formaldehyde RNA gel (Ausubel *et al.*, 1992). An ~50- $\mu$ g RNA sample was loaded on a 1% agarose gel containing 1 $\times$  MOPS buffer, 0.74% (v/v) formaldehyde, 1.9 mg/ml iodoacetamide and 0.5  $\mu$ g/ml ethidium bromide. Electrophoresis was carried out in 1 $\times$  MOPS buffer at 5 V/cm. RNA was transferred to BrightStar-Plus membrane (Ambion) using TurboBlotter (Schleicher & Schuell) and Ambion RNA transfer buffer. RNA was crosslinked to the air-dried membranes using 254 nm light (Ausubel *et al.*, 1992).

DNA probes were prepared by the random priming method (Ausubel *et al.*, 1992) using [<sup>32</sup>P]dCTP and a DNA labeling kit (Pharmacia). Hybridization was carried out for 8–16 h at 42°C in Prehybridization/Hybridization Solution (Ambion). Filters were washed according to the manufacturer's protocol and subjected to autoradiography.

#### **Gel shift assay**

PCR was used to extend the Cup9p ORF with a sequence encoding Ser-Gly-Gly-Thr-His<sub>6</sub>, yielding Cup9p-H<sub>6</sub>, and to engineer flanking restriction sites (*NdeI* and *BamHI*) for insertion into pET-11c (Novagen). Cup9p-H<sub>6</sub> was overexpressed in *E.coli* BL21 (DE3) (Novagen) (Ausubel *et al.*, 1992) and purified on a 3 ml Ni-NTA column (Qiagen), using a linear gradient of imidazole. Cup9p-H<sub>6</sub> eluted at ~0.25 M imidazole (~90% pure at this step); it was dialyzed at 4°C against 10% glycerol, 0.1 M KCl, 1 mM EDTA, 0.5 mM dithiothreitol, 20 mM HEPES, pH 7.9, and then snap-frozen in multiple samples in liquid N<sub>2</sub>, and stored at -80°C. The proximal (-1 to -447) and distal (-448 to -897) PTR2 promoter probes for the gel shift assay were constructed by PCR amplification in the presence of [ $\alpha$ -<sup>32</sup>P]dCTP, and were purified using spin columns (Qiagen). The gel shift reactions (20  $\mu$ l) contained 50  $\mu$ g/ml poly-dI-dC (Pharmacia); ~1.5  $\mu$ g/ml (500 c.p.m.) DNA probe; 1 mg/ml acetylated serum albumin (New England Biolabs) and either 0.1, 0.5 or 1  $\mu$ g/ml of Cup9p-H<sub>6</sub> in 10% glycerol, 0.1 M KCl, 2.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 20 mM HEPES, pH 7.9. The samples were incubated for 30 min at room temperature, then loaded onto a 4%

polyacrylamide gel (40:1, acrylamide:bis-acrylamide) in 0.5 $\times$  TBE (Ausubel *et al.*, 1992), and electrophoresed at 10 V/cm for 3 h at 4°C, followed by autoradiography.

#### **Pulse-chase analysis of Cup9p**

One hundred ml cultures of *S.cerevisiae* JD52 (*UBR1*) and JD55 (*ubr1Δ*) carrying either pCB210 (expressing Cup9p-FLAG from the *CUP9* promoter) or pCB200 (vector alone) were grown to an *A*<sub>600</sub> of ~1 in SD(-Leu) medium. Cells (50 *A*<sub>600</sub> units total) from each of the four cultures were gently pelleted by centrifugation, washed with 5 ml of SD(-Leu), pelleted again, resuspended in 2 ml of SD(-Leu), and incubated at 30°C for 10 min. Each sample was labeled for 5 min with 1.4 mCi of [<sup>35</sup>S]methionine/cysteine (EXPRESS, New England Nuclear) at 30°C, followed by pelleting in a microfuge for ~15 s. The cells were resuspended in 2.6 ml of SD(-Leu), 5 mM L-methionine, 5 mM L-cysteine, and incubated at 30°C. Samples of 0.5 ml were withdrawn during the incubation, pelleted and resuspended in 0.15 ml of 0.5 M NaCl-Lysis Buffer (1% Triton X-100, 0.5 M NaCl, 5 mM EDTA, 50 mM Na-HEPES, pH 7.5) containing a mixture of protease inhibitors (Ghislain *et al.*, 1996). Glass beads (0.5 mm) were added, and cells were disrupted by vortexing (six times, for 30 s each, with 1 min incubations on ice in between), followed by the adjustment of NaCl concentration to 0.15 M through the addition of 75 mM NaCl-Lysis Buffer, further vortexing for 30 s, and centrifugation at 12 000 g for 10 min. The volumes of supernatants were adjusted to equalize the amounts of 10% trichloroacetic acid-insoluble <sup>35</sup>S. Cup9p-FLAG was immunoprecipitated by the addition of 20  $\mu$ l of the monoclonal anti-FLAG M2 antibody conjugated to agarose beads (Kodak). Suspensions were incubated at 0°C for 1 h, with rotation, then centrifuged at 12 000 g for 30 s, and washed four times with 0.8 ml of 0.15 M NaCl-Lysis Buffer. The pellets were resuspended in SDS-sample buffer, heated at 100°C for 3 min, and subjected to SDS-12% PAGE, followed by autoradiography and quantitation using a PhosphorImager (Molecular Dynamics).

Pulse-chase analysis of GST-Cup9p-ha<sub>2</sub> was carried out as described by Bartel *et al.* (1990). Approximately 10 *A*<sub>600</sub> units of galactose-induced cells were labeled for 5 min with 0.3 mCi of [<sup>35</sup>S]EXPRESS in 400  $\mu$ l SM-galactose (-Ura) at 30°C. The cells were then transferred to microfuge tubes, pelleted and resuspended in 500  $\mu$ l of SD (-Ura), 5 mM L-methionine, 5 mM L-cysteine. Samples of 0.1 ml were withdrawn during the incubation, pelleted and lysed as above. The <sup>35</sup>S-labeled GST-Cup9p-ha<sub>2</sub> was purified using glutathione-agarose beads (Sigma) which had been blocked with bovine serum albumin (BSA; 10 mg/ml). Twenty  $\mu$ l of glutathione-agarose beads were added to each sample and the suspensions were incubated at 0°C for 60 min, with rotation, followed by washes and electrophoretic analyses as described for Cup9p-FLAG.

#### **Purification and N-terminal sequencing of Cup9p-FLAG**

Four 2-l cultures of JD55 (*ubr1Δ*) carrying pCB116 that expressed Cup9p-FLAG from the *GAL1* promoter were grown under selection in SM-raffinose to an *A*<sub>600</sub> of ~0.8, followed by transfer to SM-galactose and incubation at 30°C for 3 h. Longer induction times resulted in a Cup9p-mediated cytotoxicity and lower yields of Cup9p-FLAG. The cells (~1 $\times$ 10<sup>11</sup>) were harvested and lysed at 4°C as described by Burgers (1995). The extract was fractionated by precipitation with 0.4% Polymyxin P (Sigma) and then further by precipitation with 48% saturated ammonium sulfate. The pellet was dissolved in 3 ml of TBS buffer (0.15 M NaCl, 50 mM Tris-HCl, pH 7.5), and passed through Sephadex G-25 in TBS. The resulting sample (8.3 ml, ~70 mg/ml of protein) was applied to a column (1 ml) of the monoclonal anti-FLAG M2 antibody (Kodak). The column was washed three times with 10 ml of TBS, and Cup9p-FLAG was eluted by the addition of five 1-ml samples of TBS containing, respectively, 50, 100, 200 and 200  $\mu$ g/ml of the FLAG peptide (Kodak). Peak Cup9p-FLAG fractions (detected by immunoblotting) were concentrated by partial lyophilization, followed by precipitation with methanol (Wessel and Flügel, 1984) in the presence of human insulin (Sigma, 0.3 mg/ml) as a carrier. The resulting sample was fractionated by SDS-12% PAGE and electroblotted onto Pro-Blot membrane (Perkin-Elmer). After a brief staining with Coomassie, the band of the 37 kDa Cup9p-FLAG (~15 pmol) was excised from the membrane. Half of the sample was used to determine the amino acid composition; the other half was subjected to N-terminal sequencing for seven cycles, using the Applied Biosystems 476A protein sequencer at the Caltech Microchemistry Facility.

#### **Acknowledgements**

We thank R.Deshaines, R.J.Dohmen, S.Grigroryev, M.Hochstrasser, S.Jentsch, E.S.Johnson, K.Madura, I.Ota and C.Trotta for the gifts of

strains and plasmids; S.Carter for his guidance in the synthesis of Leu- $\beta$ Th; N.Johnsson, R.Deshaines, R.J.Dohmen and A.Webster for helpful discussions; L.Peck, Y.T.Kwon, A.Webster and F.Du for comments on the manuscript; S.Saha for help in overexpressing Cup9p; and N.Riley for technical assistance. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530).

## References

- Alagramam,K., Naider,F. and Becker,J.M. (1995) A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.*, **15**, 225–234.
- Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (1992) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York.
- Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell*, **56**, 1019–1032.
- Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.
- Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.
- Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.
- Balzi,E., Choder,M., Chen,W.A., Varshavsky,A. and Goffeau,A. (1990) Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**, 7464–7471.
- Bartel,B. (1990) *Molecular Genetics of the Ubiquitin System: the Ubiquitin Fusion Proteins and Proteolytic Targeting Mechanisms*. PhD Thesis, M.I.T., Cambridge, MA, USA.
- Bartel,B., Wünning,I. and Varshavsky,A. (1990) The recognition component of the N-end rule pathway. *EMBO J.*, **9**, 3179–3189.
- Burgers,P.M.J. (1995) Preparation of extracts from yeast and avoidance of proteolysis. *Methods Mol. Cell. Biol.*, **5**, 330–335.
- Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.
- Chen,P., Johnson,P., Sommer,T., Jentsch,S. and Hochstrasser,M. (1993) Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MATα2 repressor. *Cell*, **74**, 357–369.
- Christianson,T.W., Sikorski,R.S., Dante,M. and Hietter,P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.
- deGroot,R.J., Rümenapf,T., Kuhn,R.J. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.
- Dohmen,R.J., Stappen,R., McGrath,J.P., Forrová,H., Kolarov,J., Goffeau,A. and Varshavsky,A. (1995) An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. *J. Biol. Chem.*, **270**, 18099–18109.
- Finley,D., Özkanak,E. and Varshavsky,A. (1987) The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation and other stresses. *Cell*, **48**, 1035–1046.
- Ghislain,M., Dohmen,R.J., Lévy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. *EMBO J.*, **15**, 4884–4899.
- Herskowitz,I. (1989) A regulatory hierarchy for cell specialization in yeast. *Nature*, **342**, 749–757.
- Hilt,W. and Wolf,D.H. (1996) Proteasomes: destruction as a programme. *Trends Biochem. Sci.*, **21**, 96–102.
- Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.
- Hochstrasser,M. and Varshavsky,A. (1990) *In vivo* degradation of a transcriptional regulator: the yeast  $\alpha2$  repressor. *Cell*, **61**, 697–708.
- Island,M.D., Naider,F. and Becker,J.M. (1987) Regulation of dipeptide transport in *S.cerevisiae* by micromolar amino acid concentrations. *J. Bacteriol.*, **169**, 2132–2136.
- Jentsch,S. (1992) The ubiquitin-conjugating system. *Annu. Rev. Genet.*, **26**, 179–207.
- Jentsch,S. and Schlenker,S. (1995) Selective protein degradation: a journey's end within the proteasome. *Cell*, **82**, 881–884.
- Johnson,E.S., Gonda,D.K. and Varshavsky,A. (1990) *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. *Nature*, **346**, 287–291.
- Johnson,E.S., Bartel,B., Seufert,W. and Varshavsky,A. (1992) Ubiquitin as a degradation signal. *EMBO J.*, **11**, 497–505.
- Johnson,E.S., Ma,P.C.M., Ota,I.M. and Varshavsky,A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.*, **270**, 17442–17456.
- King,R.W., Deshaies,R.J., Peters,J.M. and Kirschner,M.W. (1996) How proteolysis drives the cell cycle. *Science*, **274**, 1652–1659.
- Knight,S.A.B., Tamai,K.T., Kosman,D.J. and Thiele,D.J. (1994) Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.*, **14**, 7792–7804.
- Madura,K. and Varshavsky,A. (1994) Degradation of  $\text{G}\alpha$  by the N-end rule pathway. *Science*, **265**, 1454–1458.
- Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.
- Mumberg,G., Müller,R. and Funk,M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.*, **22**, 5767–5768.
- Ooi,C.E., Rabinovich,E., Dancis,A., Bonifacino,J.S. and Klausner,R.D. (1996) Copper-dependent degradation of the *Saccharomyces cerevisiae* plasma membrane copper transporter Ctrlp in the apparent absence of endocytosis. *EMBO J.*, **15**, 3515–3523.
- Rothstein,R. (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.*, **194**, 281–301.
- Rubin,D.M., van Noort,S., Glickman,M., Cox,O., Wefes,I., Sadis,S., Fu,H., Goldberg,A., Vierstra,R. and Finley,D. (1997) ATPase and ubiquitin-binding proteins of the yeast proteasome. *Mol. Biol. Rep.*, **24**, 17–26.
- Schmitt,M.E., Brown,T.A. and Trumper,B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 3091–3092.
- Schwoeb,E., Bohm,T., Mendenhall,M.D. and Nasmyth,K. (1994) The B-type cyclin kinase inhibitor p40 (Sic1) controls the G1 to S transition in *Saccharomyces cerevisiae*. *Cell*, **79**, 233–244.
- Sherman,F., Fink,G.R. and Hicks,J.B. (1986) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sikorski,R.S. and Hietter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. *Genetics*, **122**, 19–27.
- Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, **67**, 31–38.
- Smith,R.L., Redd,M.J. and Johnson,A.D. (1995) The tetratricopeptide repeats of Ssn6 interact with the homeodomain of  $\alpha2$ . *Genes Dev.*, **9**, 2903–2910.
- Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.
- Wessel,D. and Flügge,U.I. (1984) A method for quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.*, **138**, 141–143.
- Wolberger,C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.*, **6**, 62–68.
- Zhou,P. and Thiele,D.J. (1993) Copper and gene regulation in yeast. *Biofactors*, **4**, 105–115.

Received August 13, 1997; revised October 13, 1997;  
accepted October 14, 1997

## The N-End Rule Pathway in *Xenopus* Egg Extracts

Ilia V. Davydov, Debabrata Patra, and Alexander Varshavsky<sup>1</sup>  
*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Received April 7, 1998, and in revised form June 30, 1998

**Ubiquitin-dependent degradation of intracellular proteins underlies a multitude of biological processes, including the cell cycle, cell differentiation, and responses to stress.** One ubiquitin-dependent proteolytic system is the N-end rule pathway, whose targets include proteins that bear destabilizing N-terminal residues. This pathway, which has been characterized only in somatic cells, is shown here to be present also in germ line cells such as the eggs of the amphibian *Xenopus laevis*. We demonstrate that the set of destabilizing residues in the N-end rule pathway of *Xenopus* eggs is similar, if not identical, to that of somatic cells such as mammalian reticulocytes and fibroblasts. It is also shown that the degradation of engineered N-end rule substrates in egg extracts can be strongly and selectively inhibited by dipeptides bearing destabilizing N-terminal residues. This result allowed us to ask whether selective inhibition of the N-end rule pathway in egg extracts influences the apoptosis-like changes that are observed in these extracts. A dipeptide bearing a bulky hydrophobic (type 2) destabilizing N-terminal residue was found to delay the apoptotic changes in egg extracts, whereas dipeptides bearing basic (type 1) destabilizing N-terminal residues had no effect. High activity of the N-end rule pathway in egg extracts provides an alternative to reticulocyte extracts for the *in vitro* analyses of this pathway.

© 1998 Academic Press

**Key Words:** ubiquitin; proteolysis; N-end rule; apoptosis; germ line; *Xenopus*.

A number of regulatory circuits, including those that control the cell cycle, cell differentiation, and responses to stress, involve metabolically unstable proteins (1–4). A short *in vivo* half-life of a regulator provides a way to generate its spatial gradients and allows for rapid

<sup>1</sup> To whom correspondence should be addressed at Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. Fax: (626) 440-9821. E-mail: avarsh@cco.caltech.edu.

adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis or degradation. Damaged or otherwise abnormal proteins tend to be short-lived as well.

Features of proteins that confer metabolic instability are called degradation signals, or degrons (5). The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (6, 7). The set of amino acid residues which are destabilizing in a given cell yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule pathway are present in all organisms examined, from mammals to fungi and bacteria (7).

In eukaryotes, an N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (8–10). The Lys residue is the site of formation of a multiubiquitin chain (11). The N-end rule pathway is thus a part of the ubiquitin (Ub)<sup>2</sup> system. Ub is a 76-residue eukaryotic protein whose covalent conjugation to other proteins plays a role in a multitude of biological processes (1, 12–15). In most of them, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (16–18).

It was found that linear Ub fusions are rapidly cleaved by Ub-specific proteases (UBPs) at the Ub-protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues. This technical advance led to the discovery of the N-end rule pathway (6, 7). The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by enzymatic deamidation, into the second-

<sup>2</sup> Abbreviations used: Ub, ubiquitin; DHFR, mouse dihydrofolate reductase; R-transferase, Arg-tRNA-protein transferase; CSF, cytostatic factor; TCA, trichloroacetic acid; DTT, dithiothreitol; AMC, 7-amino-methylcoumarin; UBP, Ub-specific processing protease.

ary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing residues (19–21). The primary destabilizing N-terminal residues are bound directly by N-recognin (also called E3), the recognition component of the N-end rule pathway. In *S. cerevisiae*, N-recognin is a 225-kDa protein, encoded by *UBR1*, that targets potential N-end rule substrates through the binding to their primary destabilizing N-terminal residues — Phe, Leu, Trp, Tyr, Ile, Arg, Lys, or His (7, 22). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (7).

The currently known physiological substrates of the N-end rule pathway are the RNA polymerase of Sindbis virus (23), the *GPA1*-encoded G $\alpha$  subunit of the *S. cerevisiae* heterotrimeric G protein (24), and Cup9p, a short-lived transcriptional repressor in *S. cerevisiae* that controls the expression of *PTR2*, which encodes a peptide transporter (25). The Ubr1p-Cup9p-Ptr2p circuit, which controls the import of peptides in yeast, is the first clear instance of a physiological function of the N-end rule pathway (25). Among the cells of multicellular organisms, this proteolytic system was characterized in rabbit reticulocytes (19) and L cells, a line of fibroblast-like mouse cells (26). Recently, it was suggested that the N-end rule pathway might play a role in apoptosis (programmed cell death) (7). A way to verify this conjecture would be to use a mutant that lacks the N-end rule pathway. Such mutants were constructed and characterized in *Escherichia coli* and *S. cerevisiae* (7) but not yet in multicellular organisms. Therefore we considered the use of N-end rule inhibitors.

Previous work has shown that the addition of amino acid derivatives such as dipeptides that bear destabilizing N-terminal residues to reticulocyte extract results in a strong and selective inhibition of the N-end rule pathway in the extract (19, 27). Specifically, dipeptides bearing type 1 destabilizing N-terminal residues inhibited the degradation of test N-end rule substrates bearing basic (type 1) destabilizing N-terminal residues but had no effect on the degradation of substrates bearing type 2 N-terminal residues. A converse pattern was observed with dipeptides bearing bulky hydrophobic (type 2) destabilizing N-terminal residues (19). Dipeptides added to *S. cerevisiae* cultures have been demonstrated to inhibit the N-end rule pathway *in vivo* as well (28). However, experiments with the N-end rule pathway of intact mammalian cells have shown dipeptides to be ineffective inhibitors in this setting<sup>3</sup>.

Given this latter constraint, we sought to explore the physiological consequences of inhibiting a metazoan N-end rule pathway in the inhibitor-accessible setting of a cell-free system. We also wished to determine whether the N-end rule pathway is present in germ line cells such as the eggs of the amphibian *Xenopus laevis*, a major experimental organism in studies of embryogenesis and cell cycle control (4, 29). Our findings are described below.

## MATERIALS AND METHODS

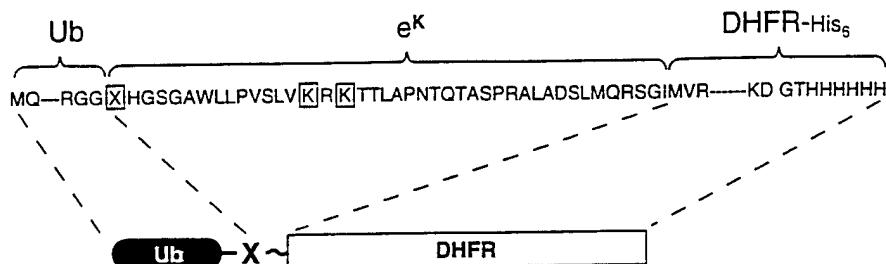
**Construction of plasmids.** The plasmids pT7-UbMe<sup>K</sup>DHFRhis and pT7-UbRe<sup>K</sup>DHFRhis expressed Ub-Met-e<sup>K</sup>-DHFR-His<sub>6</sub> and Ub-Arg-e<sup>K</sup>-DHFR-His<sub>6</sub> fusions (denoted, respectively, as Ub-Met-DHFR and Ub-Arg-DHFR) in *E. coli* from the T7 polymerase promoter (see Results for the definitions of e<sup>K</sup> and other terms). These plasmids were constructed from pEJJ1-M and pEJJ1-R (30). The *Kpn*I-*Hind*III fragment of pEJJ1-M and pEJJ1-R was replaced with a synthetic oligonucleotide duplex (5'-CCATCACCATCACCATCACTAAA-3' and 5'-AGCTTTAGTGATGGTGATGGTGATGGGTAC-3') that encoded the His<sub>6</sub> tag and bore the *Kpn*I and *Hind*III overhangs. The other pT7-UbXe<sup>K</sup>DHFRhis plasmids, which encoded otherwise identical Ub-X-DHFR fusions bearing different X residues (Met, Arg, Leu, Phe, Cys, Asp, or Asn) (Fig. 1), were constructed from pT7-UbRe<sup>K</sup>DHFRhis by site-directed mutagenesis, using PCR (31).

**Overexpression, labeling, and purification of Ub-X-e<sup>K</sup>-DHFR-His<sub>6</sub> proteins.** A pT7-UbXe<sup>K</sup>DHFRhis plasmid was introduced into *E. coli* BL21 (DE3) (31). Protein expression was induced in *E. coli* by adding isopropyl-1-thio-β-D-galactopyranoside (IPTG), and cells were labeled with [<sup>35</sup>S]methionine/cysteine (<sup>35</sup>S-Express, New England Nuclear, Boston, MA), as described (30). The labeled cells were collected by centrifugation and disrupted by sonication, and a <sup>35</sup>S-labeled Ub-X-e<sup>K</sup>-DHFR-His<sub>6</sub> (Ub-X-DHFR) test protein was purified by affinity chromatography under nondenaturing conditions, using the Ni-NTA Spin Kit (Qiagen, Chatsworth, CA), and dialyzed against 1 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 0.1 M Tris-HCl (pH 7.7), frozen rapidly, and stored at -80°C in samples that were to be thawed only once. The specific radioactivity of [<sup>35</sup>S]Ub-X-DHFR proteins was 5–10 × 10<sup>5</sup> cpm/μg.

**Degradation assays in *Xenopus* egg extracts.** Cytostatic factor (CSF)-arrested egg extracts were prepared from *Xenopus* eggs as described (32). In most experiments, cycloheximide (0.1 mg/ml) was added to the extracts to preclude reincorporation of [<sup>35</sup>S]methionine into newly made proteins. In some experiments (see the legends to figures), egg extracts were activated by the addition of 0.4 mM CaCl<sub>2</sub> 1 h before the assay. [<sup>35</sup>S]Ub-X-DHFR test proteins were added to the extract to the final concentration of ~25 μg/ml. Dipeptides were added together with bestatin (Sigma, St. Louis, MO) to the final concentrations of 10 mM and 50 μg/ml, respectively. Bestatin was added to decrease the degradation of dipeptides in the extract (27). (Control experiments (not shown) showed that the addition of bestatin alone did not significantly inhibit the degradation of test proteins by the N-end rule pathway.) The following dipeptides and other amino acid derivatives were used: Arg-β-Ala, Ala-Lys, Lys-Ala, Trp-Ala (Sigma), and Tyr-His (Bachem Science, King of Prussia, PA). Stock samples of dipeptides were 0.5 M solutions in 10 mM K-Hepes, pH 7.5.

To follow the degradation of test proteins, reaction mixtures (0.1 ml) prepared as described above were incubated at 23°C. Samples (2.5 μl) were withdrawn in duplicate for each time point. One sample was examined by SDS-12% PAGE and autoradiography, using PhosphorImager (Molecular Dynamics, Sunnyvale, CA) (33). The other sample was used to determine the relative amount (%) of <sup>35</sup>S soluble in 5% trichloroacetic acid (TCA) (30). This parameter was calculated as follows:

<sup>3</sup> F. Lévy and A. Varshavsky, unpublished data.



**FIG. 1.** The Ub-X-DHFR test proteins. These fusions, identical except for the variable residue X adjacent to the Ub moiety, were expressed and radiolabeled in *E. coli*, purified, and used as described under Materials and Methods. X was either Met, Arg, Leu, Phe, Cys, Asp, or Asn. The *E. coli* Lac repressor-derived 42-residue sequence, denoted as e<sup>K</sup> (extension (e) containing lysines (K)) was described previously (8, 30, 37). The alternative ubiquitylation sites Lys-15 and Lys-17 are boxed in the sequence of e<sup>K</sup> (8). DHFR represents DHFR-His<sub>6</sub>, the mouse DHFR whose C-terminus was extended by 8 residues that included the His<sub>6</sub> tag.

$$\text{percent TCA-soluble } ^{35}\text{S} = \frac{X}{Y} \cdot \frac{a}{a-1} \cdot 100\%,$$

where X is the amount of TCA-soluble <sup>35</sup>S (cpm); Y is the total amount of <sup>35</sup>S (cpm) in the same sample; a is the number of Met residues in a Ub-X-DHFR test protein (a = 10 for Ub-Met-DHFR; a = 9 for the other Ub-X-DHFRs). The a - 1 term above corrects for the presence of one Met residue in Ub.

The data presented in each of the experimental Figures were produced using the same extract on the same day. Unless stated otherwise, the data were consistent between at least two (usually three or more) independent experiments carried out with different preparations of the extract.

**In vitro apoptosis assays.** Apoptotic extracts from *Xenopus* eggs were prepared according to Newmeyer *et al.* (34). The difference between control and apoptotic egg extracts was in the time intervals between hormone injections used to produce the eggs. Frogs were injected with 75 units of pregnant mare serum gonadotropin (Calbiochem, San Diego, CA) 14–30 days (in contrast to 3–10 days for control extracts) before they were induced to lay eggs by injecting 800 units of human chorionic gonadotropin (Sigma). Eggs were collected, and the extract was prepared as described (32). The extracts were used within 1 h after preparation. For most assays, the extracts were arrested in interphase by the addition of 0.4 mM CaCl<sub>2</sub> and 0.1 mg/ml cycloheximide. Demembranated *Xenopus* sperm nuclei (32), at ~1000 per μl of extract, were also added at this time. Extracts treated as described were distributed into 0.1-ml samples and incubated at 23°C. Apoptotic changes in the extracts were monitored in two ways: by observing shrinkage and disintegration of the added sperm nuclei using phase contrast and fluorescent microscopy (34) and by measuring the DEVD-specific protease activity. The latter assay determined the total activity of caspases that recognize the tetrapeptide sequence DEVD (Asp-Glu-Val-Asp) (35, 36).

The DEVD-specific protease activity was measured by incubating 5 μl of extract with a quenched fluorescent substrate Ac-DEVD-AMC (50 μM) (Bachem Bioscience) in 0.1 ml of the extract buffer (15 mM MgCl<sub>2</sub>, 1 mM DTT, 20 mM EGTA, 80 mM Na-β-glycerophosphate, pH 7.3). The reaction was carried out for 10 min at 23°C, followed by freezing of a 20-μl sample in liquid N<sub>2</sub>. Samples were thawed by diluting into 1 ml of phosphate-buffered saline, and their fluorescence (excitation at 380 nm and emission at 460 nm) was measured in the Hitachi F-4500 spectrofluorimeter (35, 36). AMC was used as a fluorescence standard.

## RESULTS

### The N-End Rule Pathway in *Xenopus* Egg Extracts

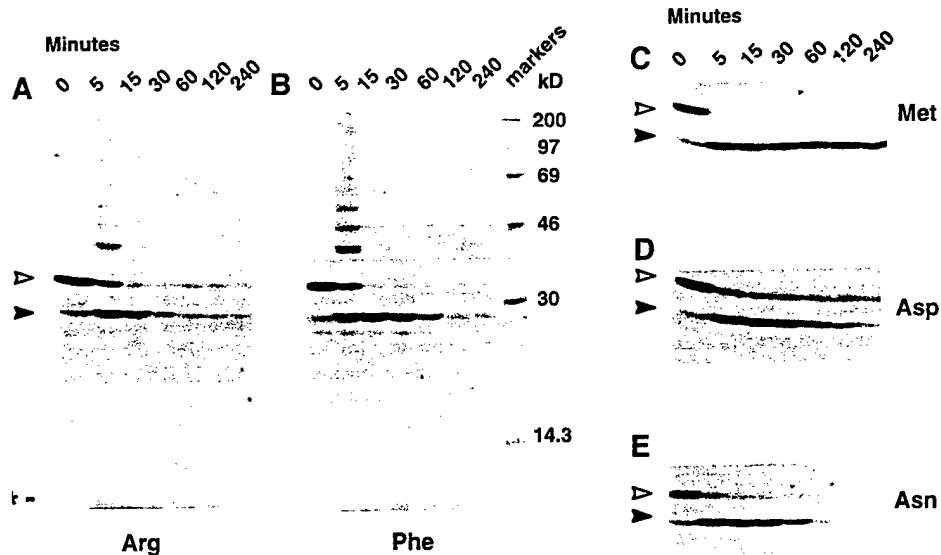
Among the cells of multicellular organisms, the N-end rule pathway has been identified and analyzed in rabbit

reticulocytes (19) and L cells, a line of transformed, fibroblast-like mouse cells (26). To determine whether the N-end rule pathway is present in germ line cells such as *Xenopus* eggs, we used an approach similar to the earlier one with rabbit reticulocytes (19, 30).

A set of N-end rule substrates differing exclusively by their N-terminal residues was constructed, expressed in *E. coli*, metabolically labeled with [<sup>35</sup>S]methionine, and purified by affinity chromatography (see Materials and Methods). A test protein of this set, Ub-X-e<sup>K</sup>-DHFR-His<sub>6</sub>, denoted below as Ub-X-DHFR, contained the following parts: the N-terminal Ub moiety; a variable residue X; e<sup>K</sup>, a 42-residue, *E. coli* Lac repressor-derived sequence that contained the second (lysine) determinant of the N-degron (8, 37); and the His<sub>6</sub>-tagged mouse dihydrofolate reductase moiety (Fig. 1). Ub fusions are not cleaved at the Ub-protein junction in *E. coli*, which lacks the Ub system (38). By contrast, in eukaryotes the Ub fusions, including the Ub-X-DHFRs, are rapidly cleaved by UBP<sub>s</sub> after the last residue of Ub, making it possible to produce, *in vivo* or *in vitro*, otherwise identical test proteins such as X-DHFRs that differ exclusively by their N-terminal residues (6, 7).

CSF-arrested *Xenopus* egg extracts were prepared from unfertilized frog eggs (32). Specific <sup>35</sup>S-labeled Ub-X-DHFR test proteins (Fig. 1) were added to the egg extracts, and their metabolic fates were monitored either by SDS-PAGE and autoradiography or by measuring the amount of TCA-soluble <sup>35</sup>S released during incubation at 23°C. The identity of residue X in the tested X-DHFRs encompassed at least one representative of each class of N-terminal destabilizing residues: Asn (tertiary destabilizing); Asp and Cys (secondary destabilizing), Arg (type 1 primary destabilizing), Phe and Leu (type 2 primary destabilizing), and Met (stabilizing) (see introduction) (7).

As expected, a Ub-X-DHFR test protein was deubiquitylated upon its addition to the extract. This was detected through an increase in the electrophoretic mobility of the major <sup>35</sup>S-labeled species and the con-



**FIG. 2.** Deubiquitylation and degradation of Ub-X-DHFR test proteins in *Xenopus* egg extract. (A)  $^{35}\text{S}$ -labeled, purified Ub-Arg-DHFR was added to a CSF-arrested *Xenopus* egg extract and incubated at 23°C for the indicated times. The samples were analyzed by SDS-PAGE, and the labeled species were detected by autoradiography (see Materials and Methods). (B) Same as A, but with Ub-Phe-DHFR. (C) Same as A, but with Ub-Met-DHFR (only the area around Met-DHFR is shown). (D) Same as C, but with Ub-Asp-DHFR. (E) Same as C, but with Ub-Asn-DHFR. Open arrowheads indicate the bands of the initial, 34-kDa Ub-X-DHFR. Closed arrowheads indicate the bands of free  $^{35}\text{S}$ Ub, which migrated close to the dye front in this electrophoretic system. The molecular masses (in kDa) of protein markers are indicated to the right of B.

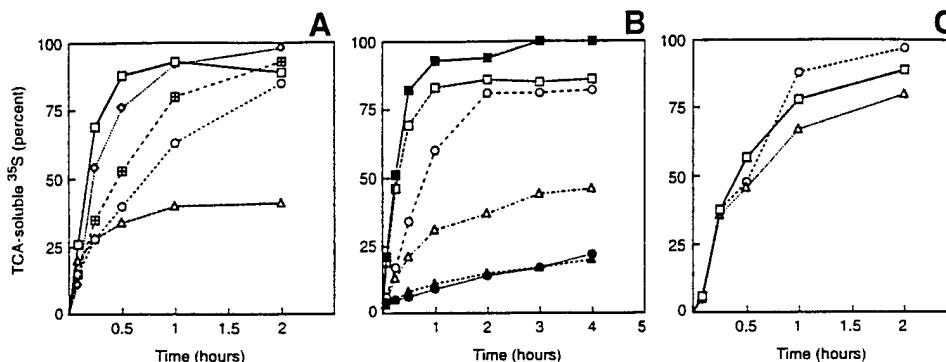
conitant appearance of a labeled ~8 kDa species (Fig. 2 and data not shown). Deubiquitylation of Ub-X-DHFR in the extract was essentially complete in 5 min or less (Fig. 2). Whereas the relative intensity of the free Ub band remained nearly constant during the subsequent 4-h incubation (Figs. 2A and 2B, and data not shown), the relative intensity of the X-DHFR band was either constant or decreased, at a varying rate, depending on the identity of the N-terminal residue X. Based on their relative metabolic stabilities in the extract (Figs. 2 and 3A, and data not shown), the X-DHFR test proteins could be reproducibly ranked from the long-lived to the short-lived as follows: Met > Cys > Leu > Asp > Asn > Phe > Arg. We conclude that *Xenopus* egg extracts contain an N-end rule pathway whose rule book is similar, and may be identical, to those found in rabbit reticulocytes (19) and mouse L cells (26).

The decay curves of X-DHFRs that were determined by monitoring the release of TCA-soluble  $^{35}\text{S}$  and the resulting ranking of their N-terminal destabilizing residues were reproducible between experiments that utilized independent preparations of the egg extract (data not shown). Specifically, it took ~10 min to destroy 50% of Arg-DHFR to acid-soluble fragments. By 30 min, the degradation of Arg-DHFR was ~90% complete (Figs. 2A and 3A).

Approximately 20% of the added Met-DHFR, which bore a stabilizing N-terminal residue (7), was degraded during the first 10 min of incubation, and another 5 to

15% was degraded over the next 30 min. However, the rest of the Met-DHFR remained stable in the extract (Figs. 2C and 3A). A likely explanation of this result is that our preparation contained two types of Met-DHFR molecules: ~70% were undamaged and long-lived in the extract, as observed with Met-DHFR *in vivo*, in both *S. cerevisiae* and mouse L cells (Ref. 8; F. Lévy and A. V., unpubl. data), whereas the remainder comprised misfolded or otherwise damaged Met-DHFR molecules that were degraded by a proteolytic system distinct from the N-end rule pathway. By inference, the same background of degradation of a subpopulation of molecules would be expected for other X-DHFRs. It is unclear whether a subpopulation of misfolded or otherwise damaged X-DHFRs resulted from their overexpression in *E. coli*, or whether the damage occurred largely during purification of X-DHFRs. Independent evidence for this explanation of the background degradation of Met-DHFR is presented below.

Previous work has shown that N-terminal Cys is a stabilizing residue in *S. cerevisiae* (8), but is a secondary destabilizing residue in rabbit reticulocyte extracts and mouse L cells (19, 26). In *Xenopus* egg extracts, Cys was found to be a weakly destabilizing residue. Specifically, among the tested X-DHFRs, Cys-DHFR was the longest-lived test protein save for Met-DHFR, which bore a stabilizing N-terminal residue (Fig. 3A). Substrates that bear secondary or tertiary destabilizing residues require the tRNA-dependent activity of R-transferase for their degradation by the N-end rule



**FIG. 3.** Degradation of X-DHFR test proteins in egg extracts under different conditions, as determined by measuring TCA-soluble <sup>35</sup>S. (A) <sup>35</sup>S-labeled, purified Ub-Arg-DHFR (□), Ub-Phe-DHFR (○), Ub-Met-DHFR (△), Ub-Asp-DHFR (○), and Ub-Asn-DHFR (■) were added to samples of the CSF-arrested *Xenopus* egg extract. After incubation at 23°C for the indicated times, the amounts of <sup>35</sup>S soluble in 5% TCA were determined for each time point (see Materials and Methods). Since the free [<sup>35</sup>S]Ub, produced through deubiquitylation of a Ub-X-DHFR, was stable during the incubation, 100% of TCA-soluble <sup>35</sup>S represented complete degradation of the Ub-lacking [<sup>35</sup>S]X-DHFR (see Materials and Methods for further details). (B) Pretreatment of egg extract with RNAase selectively inhibits degradation of N-end rule substrates bearing secondary destabilizing N-terminal residues. A CSF-arrested extract was pretreated with 5 units/ml of RNAase A-agarose (Sigma) for 1 h at 23°C, followed by the removal of RNAase-agarose by centrifugation (closed symbol curves). Another sample of the same extract was incubated for 1 h without RNAase (open symbol curves). <sup>35</sup>S-labeled Ub-Cys-DHFR (triangles), Ub-Asp-DHFR (circles), or Ub-Arg-DHFR (squares) were added to both extracts, and the relative amounts of TCA-soluble <sup>35</sup>S released during the incubation were determined as in A. (C) Different states of egg extract do not significantly alter the activity of the N-end rule pathway. <sup>35</sup>S-labeled Ub-Phe-DHFR was added to the CSF-arrested egg extract (□), to the CaCl<sub>2</sub>-treated extract that proceeded toward mitosis (△), and to the CaCl<sub>2</sub>- and cycloheximide-treated, interphase-arrested extract (○). The degradation of Phe-DHFR was monitored as described in A.

pathway (see introduction). Previous work has shown that pretreatment of reticulocyte extract with RNAase A does not significantly affect the degradation of N-end rule substrates bearing primary destabilizing residues but abolishes the degradation of otherwise identical substrates bearing secondary or tertiary destabilizing N-terminal residues (19). In agreement with these findings, preincubation of *Xenopus* egg extract with RNAase A selectively inhibited the degradation of Cys-DHFR and Asp-DHFR, but had no effect on the degradation of Arg-DHFR (Fig. 3B). We conclude that both Cys and Asp are secondary destabilizing residues in the N-end rule pathway of *Xenopus* eggs. Taken together with the earlier findings about the N-end rule pathway in reticulocyte extract (19), the present findings are consistent with the conjecture that the set of secondary and tertiary destabilizing residues in the N-end rule of *Xenopus* eggs is identical to that of rabbit reticulocytes: two tertiary destabilizing residues (Asn and Gln) and three secondary ones (Asp, Glu, and Cys).

We also asked whether the rate of degradation of an N-end rule substrate depends on physiologically relevant changes in the state of an egg extract. Unfertilized *Xenopus* eggs are arrested at the metaphase of meiosis II through the action of cytostatic factor (39, 40). Fertilization results in a transient increase of the Ca<sup>2+</sup> concentration in the eggs that inactivates CSF, allowing cells to enter interphase and proceed to mitosis. This process can be mimicked *in vitro* by the addition of Ca<sup>2+</sup> (32). Such extracts, referred to as cycling extracts, undergo changes characteristic of eggs at mi-

tosis, as observed through alterations in the morphology of *Xenopus* sperm nuclei added to the extract. The addition of cycloheximide and CaCl<sub>2</sub> to a CSF-arrested extract results not only in the escape from CSF arrest but also in a subsequent arrest at interphase, owing to the inhibition of protein synthesis, specifically the synthesis of cyclins (32). We examined the degradation of Phe-DHFR, bearing a type 2 primary destabilizing N-terminal residue, in the CSF-arrested, interphase-arrested, and cycling egg extracts, and found no significant differences in the rate of degradation of this N-end rule substrate in different extracts (Fig. 3C).

#### Dipeptides Bearing Destabilizing N-Terminal Residues Are Efficacious Inhibitors of the N-End Rule Pathway in *Xenopus* Egg Extracts

Previous work has shown that the addition of amino acid derivatives such as dipeptides bearing destabilizing N-terminal residues to reticulocyte extracts or intact yeast cells results in a strong and selective inhibition of the N-end rule pathway in these settings (19, 27). However, experiments with intact mammalian cells in culture have shown dipeptides to be at most weak inhibitors of the N-end rule pathway in this setting<sup>3</sup>. The present work stemmed in part from the possibility of exploring the physiological consequences of perturbing a metazoan N-end rule pathway in the inhibitor-accessible setting of a cell-free system such as a *Xenopus* egg extract.

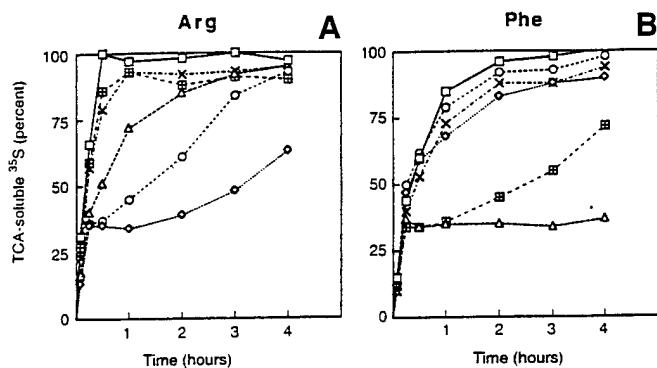


FIG. 4. Inhibition of the N-end rule pathway in egg extracts by dipeptides bearing destabilizing N-terminal residues. (A)  $^{35}\text{S}$ -labeled, purified Ub-Arg-DHFR was added to CSF-arrested *Xenopus* egg extracts in the absence (□) or the presence of the following dipeptides (10 mM), together with bestatin (50  $\mu\text{g}/\text{ml}$ ): Arg- $\beta$ -Ala (◇), Lys-Ala (○), Tyr-His (△), Trp-Ala (■), and Ala-Lys (×). After incubation at 23°C for the indicated times, the amounts of  $^{35}\text{S}$  soluble in 5% TCA were determined for each time point. (B) Same as in A but with Ub-Phe-DHFR.

As shown in Figs. 4 and 5, Lys-Ala and especially Arg- $\beta$ -Ala, both of which bear a type 1 primary destabilizing N-terminal residue, strongly inhibited the degradation of Arg-DHFR in the egg extract. This inhibition was selective, in that the same dipeptides had no significant effect on the degradation of Phe-DHFR, which bore a type 2 primary destabilizing N-terminal residue. Conversely, Tyr-His and Trp-Ala, which bear type 2 destabilizing N-terminal residues, strongly inhibited the degradation of Phe-DHFR, but had at most a small effect on the degradation of Arg-DHFR (Figs. 4 and 5). Ala-Lys, which bears a type 3 destabilizing N-terminal residue (19), did not inhibit the degradation of either Arg- or Phe-DHFRs (Fig. 4).

The dipeptides had little effect on the degradation of their cognate N-end rule substrates during the first 15 min of the incubation, but subsequent proteolysis was strongly inhibited (Fig. 4). This result is consistent with the finding that 20–30% of Met-DHFR, which bears a stabilizing N-terminal residue, was rapidly degraded, whereas the rest of the added Met-DHFR was stable (Fig. 3A). As described above, the latter finding could be accounted for if our preparation of Met-DHFR (and, by inference, of other X-DHFRs) comprised two distinct populations of molecules: undamaged ones (~70%) and hence, in the case of Met-DHFR, long-lived in the extract, and misfolded or otherwise damaged molecules that were recognized and degraded by a proteolytic system distinct from the N-end rule pathway. The failure of dipeptides to inhibit the initial burst of degradation of their cognate X-DHFRs (Fig. 4) is in agreement with this explanation.

### Apoptosis and the N-End Rule Pathway

Several groups have described the use of *Xenopus* egg extracts to analyze the process of apoptosis (34, 36, 41–43). After several hours of incubation of the interphase-arrested, apoptotic egg extracts, the activity of DEVD-specific caspases (a subgroup in the ICE/CED-3 family of cysteine proteases) (43) rises sharply, followed by fragmentation of the added sperm nuclei (34). Both the induction of caspases and the fragmentation of the added nuclei in the egg extract can be suppressed by the addition of Bcl2, a protein that inhibits apoptosis *in vivo* (34, 36). Recent work has shown that apoptotic changes in egg extract are triggered, through unknown cytosolic factors, by the release of cytochrome *c* from the mitochondrial fraction of the extract (43). Apparently Bcl2 inhibits apoptosis by blocking the release of cytochrome *c* from mitochondria (43, 44).

We wished to determine whether the apoptosis-like events in egg extracts would be perturbed by dipeptide inhibitors of the N-end rule pathway. In preliminary experiments, we found that the times of apoptotic changes in egg extracts varied considerably between independent extract preparations. Specifically, the sharp rise of the DEVD-specific protease activity and the disintegration of the added sperm nuclei could occur as early as 2 h after the start of incubation or as

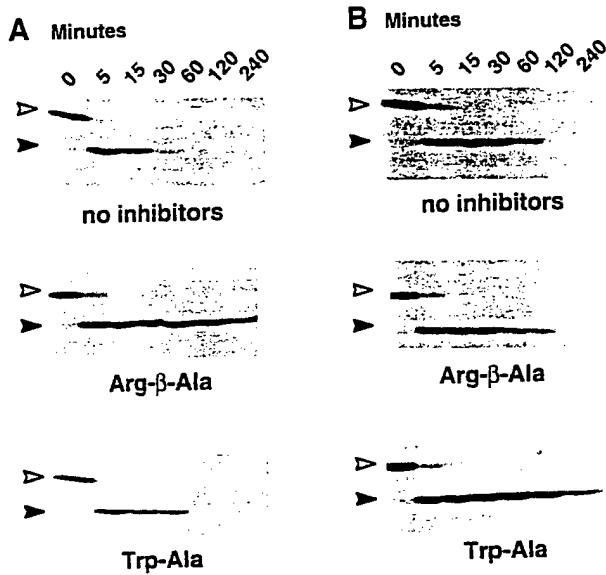


FIG. 5. Electrophoretic analysis of N-end rule substrates in egg extracts in the presence of dipeptides. (A)  $^{35}\text{S}$ -labeled, purified Ub-Arg-DHFR was added to a CSF-arrested *Xenopus* egg extract in the absence or the presence of dipeptides Arg- $\beta$ -Ala or Trp-Ala (at 10 mM, together with bestatin at 50  $\mu\text{g}/\text{ml}$ ). After incubation at 23°C for the indicated times, the samples were analyzed by SDS-PAGE, and the labeled proteins were detected by autoradiography. Open arrowheads indicate the bands of the initial, 34-kDa Ub-Arg-DHFR. Closed arrowheads indicate the bands of deubiquitylated, 26-kDa Arg-DHFR. (B) Same as in A but with Ub-Phe-DHFR.

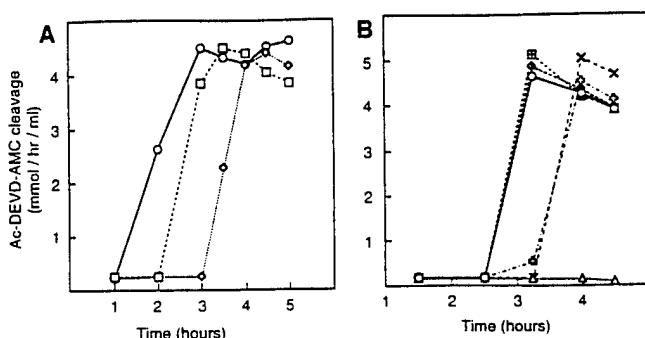


FIG. 6. Apoptotic changes in *Xenopus* egg extracts, assayed by measuring the DEVD-specific protease activity. (A) A preparation of the CSF-arrested egg extract was divided into three samples. The extracts were left untreated (□) or were activated by either 0.4 mM CaCl<sub>2</sub> (◇) or 0.4 mM CaCl<sub>2</sub> in the presence of 0.1 mg/ml cycloheximide (○). The samples were incubated at room temperature for up to 5 h. The DEVD-specific protease activity in the extracts was determined using the fluorogenic substrate Ac-DEVD-AMC. (B) Samples of interphase-arrested egg extract were incubated either in the absence (○) or in the presence of the following reagents: Lys-Ala at 20 mM (■); Ala-Lys at 20 mM (◆); Trp-Ala at 20 mM (✗); Trp-Ala plus Lys-Ala, each at 10 mM (◆); Ac-DEVD-CHO (an inhibitor of DEVD proteases, at 10 mM) (△). The dipeptides were added together with bestatin (50 µg/ml). The DEVD-specific protease activity was measured as in A.

late as 5 h (data not shown). This variability, reported by others as well (43), is likely to result in part from the variability of the timing of upstream events that trigger the beginning of apoptosis in the *Xenopus* egg extract. In addition, the timing of apoptotic changes depended on the state of the extracts: when a single preparation was split into three samples, and was used as CSF-arrested, cycling, and interphase-arrested extracts, the increase of the DEVD-specific protease activity was observed first in the interphase-arrested extract, then in the CSF-arrested one, and later in the cycling extract (Fig. 6A). This pattern was reproducible between different preparations of the extract, even though the absolute timing of the rise of the DEVD-specific protease activity varied significantly between preparations (data not shown).

We asked whether selective inhibition of the N-end rule pathway would affect the nature or kinetics of the apoptotic changes. Dipeptides that had been shown to be efficacious inhibitors of the N-end rule pathway (Figs. 2 and 3A) were added to the interphase-arrested and cycling extracts at the beginning of incubation. We found that Lys-Ala and Arg-β-Ala, containing type 1 primary destabilizing N-terminal residues, and Ala-Lys, containing a type 3 destabilizing residue, did not alter the timing of DEVD-specific protease induction (Fig. 6B). In contrast, Trp-Ala, which bears a type 2 destabilizing N-terminal residue, delayed both the rise of the DEVD-specific protease activity and the disintegration of sperm nuclei by ~1 h in 8 out of 12 independent experiments (Fig. 6B and data not shown).

The ~1 h delay of apoptotic changes by Trp-Ala was significant but weak in comparison to the effect of Ac-DEVD-CHO, an inhibitor of DEVD-specific proteases, which delayed apoptosis by at least 2 h (data not shown). Interestingly, the proteasome inhibitor 5-iodo-4-hydroxyl-3-nitrophenylacetyl-leucinyl-leucinyl-leucine-vinyl sulfone (NIP-L<sub>3</sub>VS) (45), which inhibited the degradation of Phe-DHFR in egg extracts, also delayed apoptosis in the CSF-arrested extract by ~1 h (data not shown), similarly to the effect of Trp-Ala (Fig. 6B).

## DISCUSSION

We report the following results:

(1) A whole-cell extract of *X. laevis* eggs contains the N-end rule pathway, as determined by monitoring the metabolic fates of purified, <sup>35</sup>S-labeled Ub-X-DHFR test proteins added to the extract. Among the cells of multicellular organisms, the N-end rule pathway has been identified and characterized in rabbit reticulocytes (18) and L cells, a mouse cell line (19, 26), but not, until now, in germ line cells.

(2) The partial N-end rule defined in egg extracts was found to be identical to the corresponding subsets of the N-end rules in rabbit reticulocytes and mouse fibroblasts. In particular, we showed that in egg extracts the destabilizing activity of N-terminal Asn, Asp, and Cys requires the presence of RNA (presumably tRNA), strongly suggesting that these are secondary (Asp and Cys) and tertiary (Asn) residues in the egg's N-end rule pathway, as they have been shown to be in other metazoan cells (see introduction for the terminology) (7).

(3) The activity of the N-end rule pathway did not change significantly upon shifts of the CSF-arrested egg extract to the cycling or the interphase-arrested state.

(4) Dipeptides bearing destabilizing N-terminal residues selectively and efficiently inhibited the degradation of N-end rule substrates in egg extract. In particular, dipeptides bearing the basic (type 1) primary destabilizing N-terminal residues inhibited the degradation of N-end rule substrates bearing type 1 but not type 2 N-terminal residues. A converse inhibition pattern was observed with dipeptides bearing bulky hydrophobic (type 2) N-terminal residues.

(5) The onset of apoptosis-like changes in egg extract (fragmentation of the added sperm nuclei and abrupt increase of the DEVD-specific protease activity (34, 43)) was found to be delayed by ~1 h in the presence of dipeptide Trp-Ala, which bears a type 2 destabilizing N-terminal residue, but not in the presence of dipeptides bearing type 1 destabilizing N-terminal residues.

The N-end rule pathway has been identified in all organisms examined, from mammals to fungi and bac-

teria, but the understanding of its functions remains incomplete (6, 7). The first clear example of a physiological function of the N-end rule pathway is its recently identified role in controlling the import of peptides in *S. cerevisiae* through the degradation of the transcriptional repressor Cup9p that down-regulates a peptide transporter (25, 46). It has also been reported that dipeptides bearing destabilizing N-terminal residues specifically inhibit differentiation of rat pheochromocytoma PC12 cells (47), the neurite outgrowth in amphibian neuroepithelial cells (48), and limb regeneration in the newt (49), suggesting a role for the N-end rule pathway in these biological processes. Unfortunately, none of these studies (47–49) included control experiments to verify that the added dipeptides actually inhibited degradation of a reporter N-end rule substrate in the target cells. Thus, it remains to be determined whether the observed biological effects of dipeptides (47–49) were caused by specific inhibition of the N-end rule pathway in the target cells or resulted from other effects of the added dipeptides.

Apart from the interest in determining whether the N-end rule pathway is present in germ line cells such as *Xenopus* eggs, one reason for initiating this work was the opportunity to test, through selective inhibition of the N-end rule pathway in egg extracts, whether this pathway plays a role in apoptosis. Several lines of evidence suggest a role for the Ub/proteasome system, of which the N-end rule pathway is a part, in regulating apoptosis (50–55). In particular, cell-penetrating proteasome inhibitors were reported to partially protect nonproliferating cells such as thymocytes and growth factor-deprived sympathetic neurons from apoptosis (52, 53). In contrast, the same proteasome inhibitors were reported to induce apoptosis in mitotically active cells (51, 54). An apoptosis-like process, induced by the stress of starvation and mediated by proteolysis, was also identified in *E. coli* (56). A key regulator of *E. coli* apoptosis is a short-lived protein MazE, which is degraded by ClpAP, a proteasome-like ATP-dependent protease that targets N-end rule substrates in *E. coli* (38). These and other considerations, including the possibility that cleavages by caspases can produce N-end rule substrates, led to the suggestion that the N-end rule pathway may function in the control of apoptosis (7).

Specific dipeptides, which have been shown to act as selective inhibitors of the N-end rule pathway in reticulocyte extracts (19, 27) and intact *S. cerevisiae* cells (28), were found to be largely ineffective with several lines of mammalian cells<sup>3</sup>. *Xenopus* egg extracts, in contrast to reticulocyte extracts, have been shown to undergo apoptosis-like changes (34, 36, 43), hence the choice of these extracts for the present study. We found that dipeptides bearing either type 1 (Arg, Lys) or type 3 (Ala) destabilizing N-terminal residues had no effect

on the apoptotic changes in the extracts. In contrast, Trp-Ala, which bears a type 2 destabilizing N-terminal residue, inhibited apoptosis in a number of independent experiments with different preparations of egg extracts; in a minority of tests the effect was not observed. Tyr-His, a dipeptide bearing another type 2 destabilizing N-terminal residue, strongly inhibited the degradation of an N-end rule substrate such as Phe-DHFR in the extract, but was significantly less efficacious than Trp-Ala in its effect on the timing of apoptotic changes in the same extract. Taken together, the results of experiments with dipeptides as inhibitors of apoptosis indicate that some inhibitors of the N-end rule pathway, notably Trp-Ala, can cause a significant delay of apoptosis in the extract. A definitive test and further analysis of the thus suggested function of the N-end rule pathway in apoptosis will require mutants that eliminate this pathway in a multicellular organism without perturbing the rest of the Ub system. Construction of a mouse mutant that lacks the entire N-end rule pathway (through a deletion of the *Ubr1* gene that encodes N-recognition, the main recognition component of this proteolytic system) is under way<sup>4</sup>.

In summary, we established the existence of the N-end rule pathway in germ line cells such as *Xenopus* eggs, showed that dipeptides are efficacious inhibitors of the N-end rule pathway in egg extracts, and examined the effects of dipeptides on apoptotic changes in these extracts.

#### ACKNOWLEDGMENTS

We are grateful to W. Dunphy for helpful discussions and the access to his laboratory's *Xenopus* facility, and to M. Bogoy and H. Ploegh for a gift of the NIP-L<sub>2</sub>VS proteasome inhibitor. We thank E. Smirnova and members of the Varshavsky lab, especially L. Peck, G. Turner, and A. Webster, for their suggestions and for the comments on the manuscript. D.P. is a Fellow of the Leukemia Society of America. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530).

#### REFERENCES

1. Haas, A. J., and Siepmann, T. J. (1997) *FASEB J.* **11**, 1257–1268.
2. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22**, 383–387.
3. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30**, 405–439.
4. King, R. W., Deshaies, R. J., Peters, J. M., and Kirschner, M. W. (1996) *Science* **274**, 1652–1659.
5. Varshavsky, A. (1991) *Cell* **64**, 13–15.
6. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234**, 179–186.
7. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149.
8. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56**, 1019–1032.
9. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346**, 287–291.

<sup>4</sup> Y. T. Kwon and A. Varshavsky, unpublished data.

10. Hill, C. P., Johnston, N. L., and Cohen, R. E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4136–4140.
11. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) *Science* **243**, 1576–1583.
12. Pickart, C. M. (1997) *FASEB J.* **11**, 1055–1066.
13. Wilkinson, K. D. (1997) *FASEB J.* **11**, 1245–1256.
14. Hershko, A., and Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61**, 761–807.
15. Finley, D., and Chau, V. (1991) *Annu. Rev. Cell Biol.* **7**, 25–69.
16. Cox, O., Tanaka, K., and Goldberg, A. L. (1996) *Annu. Rev. Biochem.* **65**, 801–817.
17. Hilt, W., and Wolf, D. H. (1996) *Trends Biochem. Sci.* **21**, 96–102.
18. Lupas, A., Flanagan, J. M., Tamura, T., and Baumeister, W. (1997) *Trends Biochem. Sci.* **22**, 399–404.
19. Gonda, D. K., Bachmair, A., Wünnling, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264**, 16700–16712.
20. Balzi, E., Choder, M., Chen, W., Varshavsky, A., and Goffeau, A. (1990) *J. Biol. Chem.* **265**, 7464–7471.
21. Baker, R. T., and Varshavsky, A. (1995) *J. Biol. Chem.* **270**, 12065–12074.
22. Bartel, B., Wünnling, I., and Varshavsky, A. (1990) *EMBO J.* **9**, 3179–3189.
23. deGroot, R. J., Rümenapf, T., Kuhn, R. J., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8967–8971.
24. Madura, K., and Varshavsky, A. (1994) *Science* **265**, 1454–1458.
25. Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17**, 269–277.
26. Lévy, F., Johnsson, N., Rümenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4907–4912.
27. Reiss, Y., Kaim, D., and Hershko, A. (1988) *J. Biol. Chem.* **263**, 2693–269.
28. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. USA* **87**, 2374–2378.
29. Dunphy, W. G. (1994) *Trends Cell Biology* **4**, 202–207.
30. Johnston, J. A., Johnson, E. S., Waller, P. R. H., and Varshavsky, A. (1995) *J. Biol. Chem.* **270**, 8172–8178.
31. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. Eds. (1996) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York.
32. Murray, A. W. (1991) *Methods Cell. Biol.* **36**, 581–605.
33. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. J., and Varshavsky, A. (1996) *J. Biol. Chem.* **271**, 28521–28532.
34. Newmeyer, D. D., Farschon, D. M., and Reed, J. C. (1994) *Cell* **79**, 353–364.
35. Nicholson, D. W., Ali, A., Thornberry, N. A., Vaillancourt, J. P., Ding, C. K., Gallant, M., Gareau, Y., Griffin, P. R., Labelle, M., Lazebnik, Y. A., Munday, N. A., Raju, S. M., Smulson, M. E., Yamin, T.-T., Yu, V. L., and Miller, D. K. (1995) *Nature* **376**, 37–43.
36. Cosulich, S. C., Green, S., and Clarke, P. R. (1996) *Curr. Biol.* **6**, 353–364.
37. Johnson, E. S., Ma, P. C. M., Ota, I. M., and Varshavsky, A. (1995) *J. Biol. Chem.* **270**, 17442–17456.
38. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254**, 1374–1377.
39. Sagata, N. (1996) *Trends Cell Biol.* **6**, 22–28.
40. Furuno, N., Ogawa, Y., Iwashita, J., Nakajo, N., and Sagata, N. (1997) *EMBO J.* **16**, 3860–3865.
41. Evans, E. K., Lu, W., Strum, S. L., Mayer, B. J., and Kornbluth, S. (1997) *EMBO J.* **16**, 230–241.
42. Kluck, R. M., Bossy-Wetzel, E., Green, D. R., and Newmeyer, D. D. (1997) *Science* **275**, 1132–1136.
43. Kluck, R. M., Martin, S. J., Hoffman, B. M., Zhou, J. S., Green, D. R., and Newmeyer, D. D. (1997) *EMBO J.* **16**, 4639–4649.
44. Villa, P., Kaufmann, S. H., and Earnshaw, W. C. (1997) *Trends Biochem. Sci.* **22**, 388–393.
45. Bogyo, M., McMaster, J. S., Gaczynska, M., Tortorella, D., Goldberg, A. L., and Ploegh, H. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6629–6634.
46. Alagramam, K., Naider, F., and Becker, J. M. (1995) *Mol. Microbiol.* **15**, 225–234.
47. Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **30**, 280–288.
48. Maufroid, J. P., Bradshaw, R. A., Boilly, B., and Hondermarck, H. (1996) *Int. J. Dev. Biol.* **40**, 609–611.
49. Taban, C. H., Hondermarck, H., Bradshaw, R. A., and Boilly, B. (1996) *Experientia* **52**, 865–870.
50. Delic, J., Morange, M., and Magdelenat, H. (1993) *Mol. Cell. Biol.* **13**, 4875–4883.
51. Imajoh-Ohmi, S., Kawaguchi, T., Sugiyama, S., Tanaka, K., Omura, S., and Kikuchi, H. (1995) *Biochem. Biophys. Res. Commun.* **217**, 1070–1077.
52. Grimm, L. M., Goldberg, A. L., Poirier, G. G., Schwartz, L. M., and Osborne, B. A. (1996) *EMBO J.* **15**, 3835–3844.
53. Sadoul, R., Fernandez, P.-A., Quiquerez, A.-L., Martinou, I., Maki, M., Schröter, M., Becherer, J. D., Irmler, M., Tschoop, J., and Martinou, J.-C. (1996) *EMBO J.* **15**, 3845–3852.
54. Drexler, H. C. A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 855–860.
55. Lopes, U. G., Erhardt, P., Yao, R., and Cooper, G. M. *J. Biol. Chem.* **272**, 12893–12896.
56. Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6059–6063.

# Ump1p Is Required for Proper Maturation of the 20S Proteasome and Becomes Its Substrate upon Completion of the Assembly

Paula C. Ramos,\* Jörg Höckendorff,\*  
Erica S. Johnson,†‡ Alexander Varshavsky,†  
and R. Jürgen Dohmen§¶

\*Biotechnologisches Zentraleinheit  
Institut für Mikrobiologie  
Heinrich-Heine-Universität Düsseldorf  
Universitätsstr. 1, Geb. 25.02  
40225 Düsseldorf  
Germany  
†Division of Biology  
California Institute of Technology  
Pasadena, California 91125  
‡Laboratory of Cell Biology  
The Rockefeller University  
New York, New York 10021

## Summary

We report the discovery of a short-lived chaperone that is required for the correct maturation of the eukaryotic 20S proteasome and is destroyed at a specific stage of the assembly process. The *S. cerevisiae* Ump1p protein is a component of proteasome precursor complexes containing unprocessed β subunits but is not detected in the mature 20S proteasome. Upon the association of two precursor complexes, Ump1p is encased and is rapidly degraded after the proteolytic sites in the interior of the nascent proteasome are activated. Cells lacking Ump1p exhibit a lack of coordination between the processing of β subunits and proteasome assembly, resulting in functionally impaired proteasomes. We also show that the propeptide of the Pre2p/Doa3p β subunit is required for Ump1p's function in proteasome maturation.

## Introduction

Ubiquitin (Ub)-dependent proteolysis underlies the bulk of nonlysosomal protein degradation in eukaryotic cells (reviewed by Hochstrasser, 1996; Varshavsky, 1997). Naturally short-lived as well as damaged or otherwise abnormal proteins are recognized by the Ub system and are marked for degradation by the attachment of multi-Ub chains. Ubiquitylated proteins are degraded by the 26S proteasome, an ~2000 kDa, multisubunit, ATP-dependent protease that consists of the 19S complex, which is required specifically for the degradation of ubiquitylated proteins, and an ~700 kDa complex, called the 20S proteasome, which is the ATP-independent catalytic core of the 26S proteasome (reviewed by Peters, 1994; Coux et al., 1996; Lupas et al. 1997).

The 20S proteasome is universal among eukaryotes; its structural homologs have also been found in archaeons and eubacteria (reviewed by Coux et al., 1996). High-resolution crystal structures have been reported for the

20S proteasomes of the archaeon *Thermoplasma acidophilum* and the eukaryote *Saccharomyces cerevisiae* (Löwe et al., 1995; Groll et al., 1997). The *Thermoplasma* proteasome contains two types of subunits, α and β, which form a hollow cylinder composed of four heptameric rings in the configuration α<sub>7</sub>β<sub>7</sub>α<sub>7</sub>. The yeast 20S proteasome has a similar structure but contains 14 distinct subunits, seven of the α type and seven of the β type. The N-terminal threonines of the β subunits of the *Thermoplasma* proteasome act as nucleophiles in catalyzing the hydrolysis of peptide bonds of polypeptide substrates (Seemüller et al., 1995). Lys-33 and Glu-17 of the β subunit also play a role in catalyzing the cleavage of peptide bonds. The 14 identical β subunits of the *Thermoplasma* proteasome form 14 identical active sites, which catalyze the cleavage of substrates after hydrophobic amino acid residues (chymotrypsin-like active sites). These sites are located on the inner surface of a central chamber formed by the two rings of β subunits (Löwe et al., 1995; Seemüller et al., 1995). β subunits are synthesized as inactive precursors containing a propeptide that is thought to be cleaved off autocatalytically, yielding the mature β subunit bearing N-terminal Thr. The propeptides of β subunits are not required for the in vitro assembly of the *Thermoplasma* proteasome (Seemüller et al., 1996).

In eukaryotic proteasomes, only 3 of the 7 distinct β subunits contain the three conserved residues required for activity of the *Thermoplasma* proteasome. Genetic and structural data suggest that these three subunits provide the active-site nucleophiles for the three distinct catalytic activities of eukaryotic proteasomes, namely the "chymotrypsin-like" (see above), the "trypsin-like" (cleavage after basic residues), and the "peptidylglutamyl peptide hydrolyzing (PGPH)" activity (cleavage after acidic residues). In *S. cerevisiae*, these three β subunits are Pre2p/Doa3p, Pup1p, and Pre3p (Heinemeyer et al., 1993; Chen and Hochstrasser, 1996; Arendt and Hochstrasser, 1997; Groll et al., 1997; Heinemeyer et al., 1997). In the human 20S proteasome, the related proteins are MB1, δ, and Z; during the immune response, these subunits can be replaced by their respective homologs LMP7, LMP2, and MECL-1/LMP10 (reviewed by Coux et al., 1996). In the case of LMP2 and LMP7, it has been demonstrated that these replacements are functionally relevant in altering the specificity of antigen presentation by the MHC class I pathway. The catalytically active β subunits of the eukaryotic proteasome are synthesized with propeptides, similarly to the β subunit in *Thermoplasma* (Schmidtke et al., 1997).

In the crystal structure of the *S. cerevisiae* 20S proteasome, the opening to the proteasome's interior, formed by the outer ring of the α subunits, is not large enough to admit even an unfolded polypeptide chain, let alone a folded protein (Groll et al., 1997). The degradation of larger substrates requires the 19S complexes, which attach at both sides of the 20S proteasome, yielding the 26S proteasome. The 19S complexes contain subunits that bind multi-Ub chains, at least one Ub-specific isopeptidase that disassembles these chains, and several

§ To whom correspondence should be addressed.

ATPases that are thought to be involved in perturbing the substrate conformationally and guiding it to the interior of the 20S proteasome (Deveraux et al., 1994; Hochstrasser et al., 1995; Jentsch and Schlenker, 1995; Coux et al., 1996).

Studies with mammalian cells have shown that the 20S proteasome is assembled through a 15–16S intermediate, apparently a half-proteasome. This intermediate contains all 14  $\alpha$  and  $\beta$  subunits, some of which are in the precursor (propeptide-bearing) form, and several uncharacterized polypeptides as well (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997). Studies on the yeast 20S proteasome have demonstrated that processing of proPre2p (identical to proDoa3p) is coupled to formation of the 20S proteasome from two half-proteasome precursors (Chen and Hochstrasser, 1996). Formation of the active site capable of autocatalytic processing of proPre2p depends on the juxtaposition of proPre2p and Pre1p on the opposite sides of the two halves of the proteasome. This mechanism is thought to prevent activation of proteolytic sites before the central hydrolytic chamber has been sealed off from the cytosol through association of the two halves of the proteasome (Chen and Hochstrasser, 1996). The Pre2p propeptide is essential for the formation of functional proteasomes. Moreover, this propeptide can operate in *trans*, suggesting that it serves a chaperone-like function in proteasome maturation (Chen and Hochstrasser, 1996).

In the present work, we identify Ump1p, a novel protein, as a component of a precursor complex of the 20S proteasome. This precursor contains unprocessed  $\beta$  subunits. Upon formation of the 20S proteasome from two such precursors, the propeptides of  $\beta$  subunits are removed, a process that is accompanied by removal of Ump1p from the proteasome through Ump1p degradation, which requires the proteasome's proteolytic activity. In *ump1Δ* cells, which lack Ump1p, coordination of 20S complex formation and processing of  $\beta$  subunits is impaired, resulting in incompletely or (in the case of Pre2p) prematurely processed  $\beta$  subunits. These findings reveal Ump1p as a novel type of molecular chaperone, a short-lived maturation factor required for the efficient biogenesis of the 20S proteasome. We also describe genetic evidence that the propeptide of proPre2p is required for Ump1p-dependent proteasome maturation, and we present a model that accounts for some of the functions of Ump1p and propeptides in proteasome maturation.

## Results

### The *UMP1* Gene Is Required for Ubiquitin-Mediated Proteolysis

To identify the genes required for Ub-dependent proteolysis in *S. cerevisiae*, we screened for mutants defective in the degradation of test substrates. One such mutant, termed *ump1-1* (ub-mediated proteolysis), exhibited defects in the degradation of several normally short-lived proteins (see below). The complementing *UMP1* gene encoded a 148-residue (16.8 kDa) protein. Searches of the databases did not identify close homologs of Ump1p but did detect similarities to regions of

several proteins. One intriguing similarity was between Ump1p and C-terminal regions of the protease inhibitor contrapsin (25% identity, 52% similarity) and related proteins (Figure 1A), consistent with the likely role of Ump1p as an inhibitor of premature proteolytic activation of the proteasome precursor complexes (see below).

Pulse-chase analyses in *ump1Δ* mutants revealed a strong stabilization of several normally short-lived test proteins, in particular an N-end rule substrate Arg- $\beta$ -galactosidase (Arg- $\beta$ -gal), which bears N-terminal Arg, a destabilizing residue (Varshavsky, 1996), Ub-Pro- $\beta$ -gal (a substrate of the UFD pathway; Johnson et al., 1995; Varshavsky, 1997), and  $\alpha$ 2<sub>deg1</sub>- $\beta$ -gal (a substrate of the DOA pathway; Hochstrasser and Varshavsky, 1990; Hochstrasser et al., 1995) (Figure 1B). The recognition and ubiquitylation of these substrates involve different recognins (E3 proteins) and different Ub-conjugating (E2) enzymes. The pleiotropic character of the *ump1Δ* phenotype suggested that Ump1p functions downstream of the recognition and ubiquitylation components of the Ub system, perhaps at the step of proteolysis, or in regulating the supply of Ub.

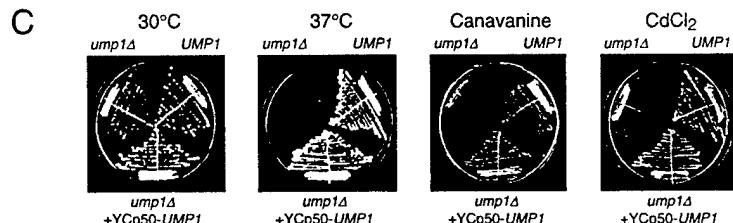
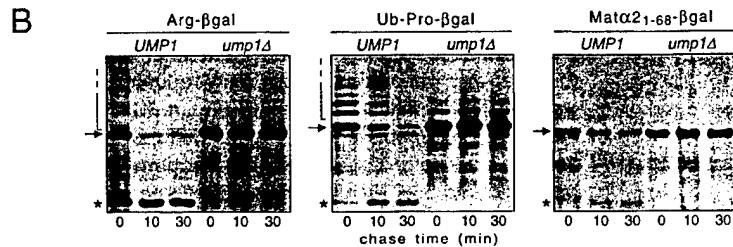
### *ump1Δ* Mutants Are Hypersensitive to a Variety of Stresses, Accumulate Ub-Protein Conjugates, and Do Not Sporulate

*ump1Δ* mutants grew more slowly than congenic wild-type (wt) cells at 30°C or lower temperatures and were severely growth-impaired at higher temperatures (37°C) (Figure 1C and data not shown). In addition, they were hypersensitive to cadmium ions and to the arginine analog canavanine (Figure 1C). Similar phenotypes have been reported for mutants in genes encoding E2 (Ubc) enzymes or proteasome subunits (Heinemeyer et al., 1993; Jungmann et al., 1993). Comparisons of proteins from whole-cell extracts of wt and *ump1Δ* cells by immunoblotting with anti-Ub antibodies showed a dramatic accumulation of Ub-protein conjugates in the *ump1Δ* mutant. At the same time, the level of free Ub was reduced in *ump1Δ* cells (data not shown). Thus, the primary cause of the *ump1Δ* phenotype appears to be the impaired ability of *ump1Δ* cells to degrade Ub-protein conjugates. Homozygous *ump1Δ* diploids (strain JD61) were unable to sporulate (data not shown).

### Ump1p Is a Component of Proteasome Precursors

To investigate Ump1p biochemically, an epitope tag was linked to its C terminus. A single copy of the modified *UMP1* gene (*UMP1-ha*), expressed from its natural promoter and chromosomal location, restored wt growth rates. We analyzed whole-cell extracts of a strain expressing Ump1p-ha and Pre1p-ha (the latter an ha-tagged  $\beta$  subunit of the proteasome) by gel filtration on Superose-6. Ump1p-ha eluted in fractions 22 and 23, both of which also contained Pre1p-ha and other proteasomal subunits (Figures 2A, 2B, and 3A) but lacked the chymotrypsin-like activity of the mature 20S proteasome. The apparent size of the Ump1p-containing complex was 300–400 kDa (Figure 2A). 20S and 26S proteasomes, as well as free 19S caps, and 20S proteasomes with one attached 19S cap, eluted in earlier fractions

A	Ump1p	1	M N I V P Q D T L K S O V S T D O D R S * L	22
	contrapsin	218	I S F D P Q D M F E S E F Y L D E K R I S M K V P P M M K M K I L T T I R H F R D E E L S C S	261
	a1act	219	M P F D P D O D T H O S R F Y L S K I K M V M P M M S L H H L T I P Y F R D E E L S C T	262
	spi-2	210	V P F D M D Y S F E S E F Y V D E K R I S V K V I S M M K I E E L T T P Y F R D E E L S C S	253
	Ump1p	23	S S A V P S * P D T L I P D E G G A V P R S T O N D R H P L	53
	contrapsin	262	V L E L K Y T G N A S A L L I T P D D G S M Q O V E N S L O P E T I R K W R P K T D F P S	305
	a1act	263	V V E L K Y T G N A S A L F I P D D K M E E V E A M I P E T I R K W R P D S L E F R	306
	spi-2	254	V L E L K Y T G N A S A L F I P D D K M E E V E A M I P E T I R K W R P D S L R P R	297
	Ump1p	54	- E S T D K N W E T T Q R O P R O M E O Y R O I F G A P M K R T M E I V N R I	94
	contrapsin	306	- Q T E E M E P F K S I A S N Y E E D V P L P E M G A K E V F T I E Q A D L S G I	347
	a1act	307	E I G E L Y I P P K F S I S R D Y N I N D - T I I L O L G S I E A F T I S K I A D L S G I	347
	spi-2	298	K I D E L Y I P P R L I S I N D - T I I L O L G S I E A F T I S K I A D L S G I	338
	Ump1p	95	D F N P I E S T N G S ! P R D I T L N K E C S I D W E D V Y P G T Q L Q A S T M V G D D V	138
	contrapsin	348	E T K K P E S V S Q V V H K A V L D V A E T G T E A A A A T G V I G G I R K A I L -	388
	a1act	348	G A R N I L A V S Q V V H K V V S D V F E E G T E A S A A T A V K I T T L S A L V E T R T	391
	spi-2	333	G A K D D S V S Q V V H K V V D V N E T G T E A A A A T G A N L V P R S I G R P -	388
	Ump1p	139	H S K I E K O L G I	148
	contrapsin	389	A V H F N R P F L F V I Y H T S A O S I L F M A K V N P N P K	418
	a1act	392	I V F R F N R P F L M I I V P T D T Q N [P] F M S K V T N P S K P R A C I K Q W G S Q	433
	spi-2	381	L L V W F N R P F L H A V S H T H G Q M I L F M A K V I N P V G A	412



(15–21), as determined by nondenaturing gel electrophoresis (Figure 2B and data not shown), by assays of the proteasome's proteolytic activity, and by immunological detection of Cim3p, a subunit specific for the 19S cap of the 26S proteasome (Ghislain et al., 1993).

Cochromatography of Ump1p-ha and Pre1p-ha in fractions 22 and 23 (Figure 2A) suggested that Ump1p is a component of proteasome precursor complexes. Consistent with this possibility, the Ump1p-containing complex had a higher mobility than the 20S proteasome upon nondenaturing gel electrophoresis. A proteasome precursor with an electrophoretic mobility indistinguishable from that of the Ump1p-containing complex was also detected in extracts from a strain that expressed Pup1p-ha, another  $\beta$  subunit of the 20S proteasome (Figure 2B).

To produce independent evidence bearing on the nature of the Ump1p-containing complex, we constructed a strain that expressed both Ump1p-ha and doubly tagged Pre1p-Flag-His<sub>6</sub> (Pre1p-FH). Affinity chromatography on Ni-NTA-agarose and anti-Flag antibody resin was then used to purify complexes containing Pre1p-FH. Ump1p-ha cofractionated with Pre1p-FH in both affinity purification steps (Figures 2C and 2D), confirming that Ump1p is a component of a distinct proteasome-related complex. This complex sedimented at ~15S in sucrose gradients (data not shown). Taken together,

Figure 1. Ump1p Is Required for Ubiquitin-Mediated Proteolysis

(A) Sequence similarities between Ump1p and C-terminal parts of mouse contrapsin, human alpha1 antitrypsin inhibitor (a1act) and mouse spi-2 (PIR accession numbers, respectively, JX0129, A90475, and S31305). The sequences were aligned using the PileUp program (GCG package, version 7.2, Genetics Computer Group, Madison, WI). Gaps (indicated by hyphens) were used to maximize alignments. Residues identical between Ump1p and at least one of the other proteins are shaded in gray. Residues identical among at least two proteins other than Ump1p are boxed.

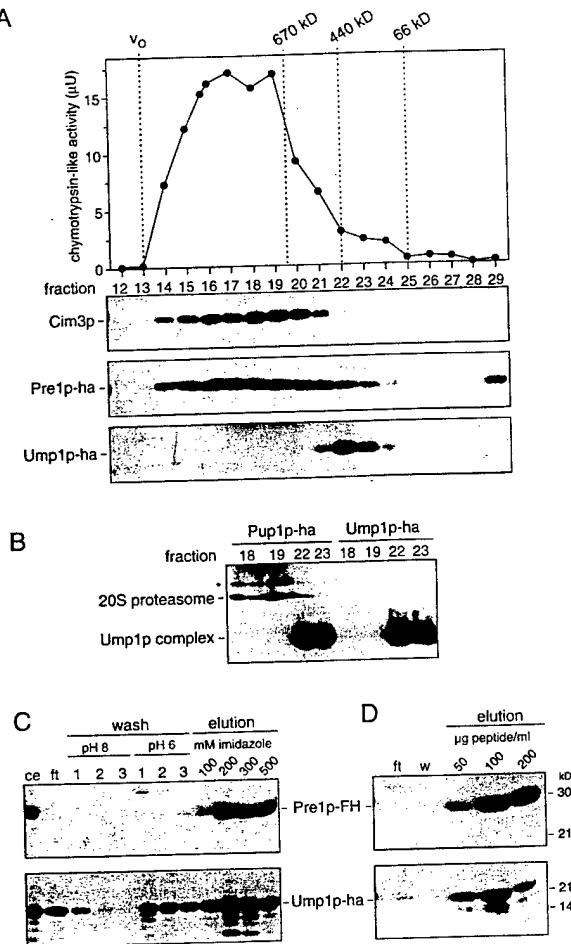
(B) Pulse-chase analysis comparing the metabolic stabilities of R- $\beta$ -gal, Ub-P- $\beta$ -gal, and MAT $\alpha$ 2-68- $\beta$ -gal in wt (*UMP1*) and *ump1Δ* cells. The open-ended brackets denote the position of multiubiquitylated  $\beta$ -gal species. Asterisks denote an ~90 kDa  $\beta$ -gal cleavage product characteristic of short-lived  $\beta$ -gal derivatives (Dohmen et al., 1991).

(C) Growth of *ump1Δ* (JD59), *UMP1* (JD47-13C), and *ump1Δ* cells transformed with YCp50-*UMP1* (expressing wild-type *UMP1*). Cells were streaked on YPD plates and incubated for 2 days at 30°C (or at 37°C where indicated), with 0.8  $\mu$ g/ml canavanine or 30  $\mu$ M CdCl<sub>2</sub> where indicated.

this evidence strongly suggested that the Ump1p-containing complex is a precursor of the 20S proteasome that is similar to the 15–16S precursors observed in the maturation pathway of mammalian proteasomes (Frentzel et al., 1994; Yang et al., 1995).

#### The Ump1p Proteasome Precursor Complex Contains Unprocessed $\beta$ Subunits

Several subunits of the proteasome are synthesized as precursors containing N-terminal extensions (propeptides) that are absent from the mature proteasome. These precursors are detected in the mammalian 15–16S complex (Frentzel et al., 1994; Yang et al., 1995). To compare the Ump1p-containing complex with the mammalian 15–16S complex, we analyzed the former for the presence of propeptide-containing proteasomal subunits. These  $\beta$  subunits were modified by the addition of C-terminal ha tags. Strains expressing Pup1p-ha, Pre2p-ha, and Pre3p-ha grew at wt rates (data not shown). When extracts of these strains were fractionated by gel filtration on Superose-6, the fractions containing the Ump1p complex contained largely the  $\beta$  subunit precursors, proPup1p-ha, proPre2p-ha, and proPre3p-ha (Figure 3A). The corresponding mature  $\beta$  subunits largely eluted in the earlier fractions that contained 20S and 26S proteasomes (Figure 3A). Subtle but reproducible differences



**Figure 2.** Ump1p Is a Component of Proteasome Precursor Complexes

(A) An extract from strain JD127 expressing Ump1p-ha and Pre1p-ha instead of wt versions of these proteins was fractionated by gel filtration on Superose-6. The upper panel shows the results of measurements of chymotrypsin-like activity in the relevant fractions (12–29). The last fraction, representing the void volume ( $v_0$ ), and positions of the peaks of marker proteins are indicated by dashed lines. The same fractions were analyzed by SDS-PAGE and immunoblotting with anti-Cim3p and anti-ha antibodies, which detected Cim3p, Pre1p-ha, and Ump1p-ha, as indicated.

(B) Analysis of Superose-fractionated 20S proteasomes and precursor complexes by nondenaturing gel electrophoresis and immunoblotting with anti-ha antibody. The strains were JD139 and JD129 expressing, respectively, Pup1p-ha and Ump1p-ha. Asterisk, 20S proteasome with one attached 19S regulator cap. The 26S proteasome eluted in fractions 14–17 of Superose-6 (not shown).

(C) Affinity purification of Pre1p-FH on Ni(NTA)-Sepharose and copurification of Ump1p-ha. Purification of proteins from an extract of strain JD126 expressing Ump1p-ha and Pre1p-FH was carried out as described in the Experimental Procedures. Different fractions were analyzed by SDS-PAGE and immunoblotting with anti-Flag (upper panel) and anti-ha antibodies (lower panel). Ce, crude extract; ft, flow-through. In contrast, no binding of Ump1p to the resin was observed when extracts from strain JD129 expressing untagged Pre1p were used (not shown).

(D) Affinity purification of Pre1p-FH on anti-Flag antibody agarose resin. Material that was eluted from the Ni(NTA)-Sepharose column shown in panel C was subjected to a second affinity purification on anti-Flag antibody resin. Pre1p-FH was specifically eluted with increasing concentrations of the Flag peptide. W, wash.

were detected among the tagged  $\beta$  subunits with respect to the distribution of their precursor and processed forms in various fractions. For example, pro-Pup1p-ha was detectable only in fractions 22 and 23, and was absent from fraction 21. By contrast, proPre2p-ha was detectable in the larger complexes (down to fraction 19), while fractions 22 and 23 already contained some processed Pre3p-ha. These patterns suggested a defined order of processing events. One conclusion from these experiments is that the Ump1p-containing complex is a precursor of the 20S proteasome that contains unprocessed  $\beta$  subunits and is therefore proteolytically inactive (Figures 3A and 3C).

#### Ump1p Is Required for Correct Proteasome Maturation

Next, we asked what effect the *ump1Δ* mutation has on maturation and activity of the proteasome. Specifically, the analyses described above were repeated with extracts obtained from *ump1Δ* cells. Figure 3C shows that there was a significant reduction in the three proteolytic activities of the proteasome in the fractions containing the 20S and 26S proteasomes (fractions 14–21). The reduction of the specific activity appears to be partially compensated by an increased expression of proteasomes (Figures 3A and 3B, data not shown) similar to that observed in cells expressing mutant  $\beta$  subunits (Arendt and Hochstrasser, 1997). In addition, a new (absent from wt cells) peak of chymotryptic activity encompassing fractions 22 and 23 was detected in extracts from *ump1Δ* cells. These were the fractions that contained the Ump1p complex from wt extracts (see above and Figure 3A), suggesting that the absence of Ump1p from the proteasome precursor complex was the cause for premature activation of its chymotryptic activity in the *ump1Δ* mutant. To verify that the detected activity was actually of the ~15S precursor complex from *ump1Δ* cells, the fraction 22 samples from strains expressing either wt Pre2p or Pre2p-ha were incubated with anti-ha antibody and protein A Sepharose. This treatment did not deplete the chymotryptic activity from the sample containing untagged Pre2p but did deplete ~80% of the activity from an otherwise identical sample derived from *ump1Δ* cells expressing Pre2p-ha (data not shown). Thus, the chymotryptic activity of this fraction resided in the proteasome precursor complexes.

To follow the appearance of the precursor and mature forms of  $\beta$  subunits in proteasome precursors and mature proteasomes, we analyzed extracts from *ump1Δ* cells expressing ha-tagged versions of these subunits by immunoblotting with anti-ha antibody. The processing of the three analyzed  $\beta$  subunits, Pup1p-ha, Pre2p-ha, and Pre3p-ha, was strikingly different in *ump1Δ* and wt cells (Figure 3A). For all three subunits, a dramatic increase of their precursors was detected in the fractions (14–21) that contained the 20S and 26S forms of the proteasome. In addition, different processed variants, possibly representing processing intermediates of pro-Pre2p, could be detected in fractions 22 and 23, which contained the ~15S proteasome precursor, and in some of the proteasome-containing fractions as well (Figure 3A). In contrast, with the exception of Pre3p, almost no processed  $\beta$  subunits were present in the fractions

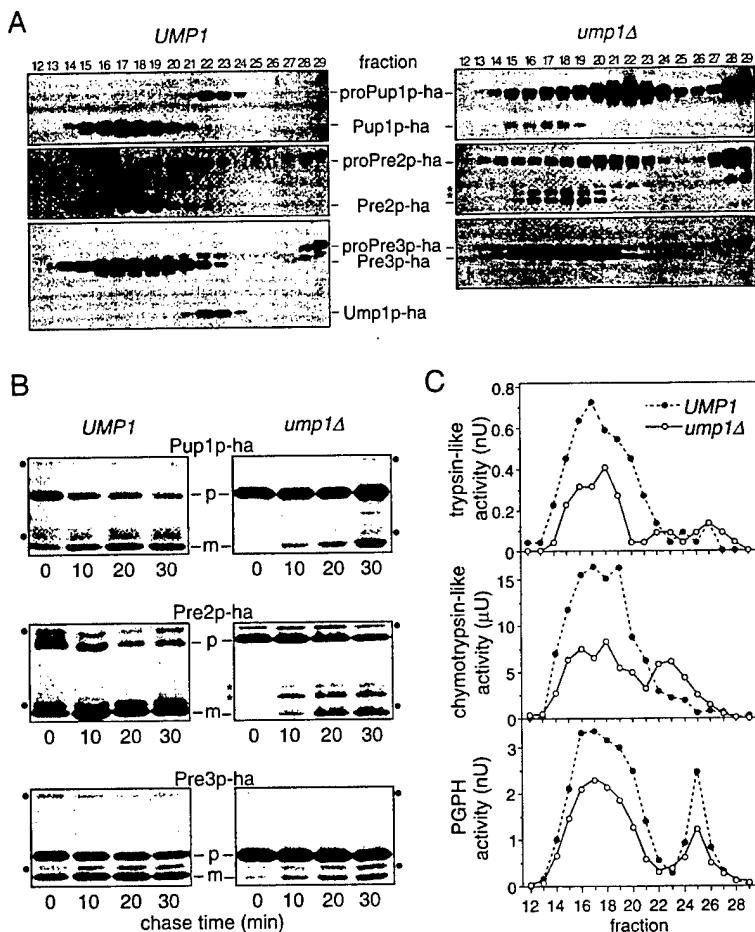


Figure 3. Ump1p Is Required for the Correct Processing of Proteasomal  $\beta$  Subunits

(A) SDS-PAGE and anti-ha immunoblot analyses of extracts fractionated on Superose-6. The congenic yeast strains used were JD131, JD132, JD133, JD134, JD135, and JD136 (Table 1). Western blots of the *ump1 $\Delta$*  extracts were developed for a shorter time to compensate for an increased expression of proteasomal subunits in the mutant. Asterisks, processing intermediates or incompletely processed forms of Pre2p-ha that were observed in the *ump1 $\Delta$*  mutant but not in wt cells.

(B) Pulse-chase analysis comparing the rate of  $\beta$  subunit processing in wt cells to that in *ump1 $\Delta$*  cells (same strains as in A). Cells were pulse-labeled with  $^{35}$ S-Met/Cys for 5 min. Proteins were precipitated with anti-ha antibody. Fluorographic exposures were ~3 times longer for the *UMP1* samples. P, propeptide-containing precursor form; m, mature form; asterisks, as in (A); dots, nonspecific bands present in some samples.

(C) Comparison of proteolytic activities in extracts with equal amounts of total protein of *UMP1* (JD127) and *ump1 $\Delta$*  cells (JD75) fractionated by gel filtration on Superose-6. The profile of the chymotrypsin-like activity in *UMP1* is the same as in Figure 2A.

containing the ~15S proteasome precursor from wt cells, and no processing intermediates of Pre2p could be detected either (Figure 3A). These findings indicate that Ump1p is required to prevent premature processing of at least proPre2p and that the presence of Ump1p is important for the coordination of proteasome assembly and subunit processing.

A defect of *ump1 $\Delta$*  cells in the processing of  $\beta$  subunits was also observed by following the metabolic fate of these subunits in pulse-chase experiments. In wt cells, most of the propeptide-containing forms of  $\beta$  subunits (proPup1p-ha, proPre2p-ha and proPre3p-ha) were converted into their mature counterparts during the 30-min chase (Figure 3B). By contrast, in the *ump1 $\Delta$*  mutant, the bulk of the  $\beta$  subunit precursors remained unprocessed during this time, a finding consistent with the results of immunoblot analyses (Figure 3A). In the case of Pre2p-ha, pulse-chase analysis again revealed the species of intermediate size (putative processing intermediates) in extracts from *ump1 $\Delta$*  cells. These species were absent from the equivalent samples derived from wt cells (Figure 3B).

#### Ump1p Is Degraded upon Formation of the 20S Proteasome

The experiments above demonstrated that Ump1p is a component of proteasome precursors that is absent from the mature 20S proteasome. One possibility was

that Ump1p leaves the precursor complex upon formation of the 20S proteasome from the two ~15S precursors and may then assist with another round of proteasome assembly, thus acting catalytically. It was also possible that formation of the 20S form of the proteasome sterically traps Ump1p. Since the formation of the 20S form of the proteasome coincides with the appearance of its proteolytic activities (Frentzel et al., 1994; Chen and Hochstrasser, 1996), Ump1p might be degraded by the newly formed 20S proteasome. Pulse-chase analysis of Ump1p-ha in wt cells showed that Ump1p-ha is indeed degraded in vivo, the rate of its degradation being similar to the rate of disappearance of proPup1p-ha, which is converted to Pup1p-ha upon maturation of the proteasome (compare Figures 4A and 3B).

#### Ump1p Is Stabilized in Proteasome Mutants and Can Be Detected Inside the 20S Particle

If Ump1p becomes a substrate of the newly formed 20S proteasome, the degradation of Ump1p should be inhibited by mutations that affect the proteasome's proteolytic activities. We used pulse-chase assays to follow the metabolic fate of Ump1p-ha in the *pre1-1* mutant, which is known to be deficient in the chymotryptic activity of the proteasome (Heinemeyer et al., 1993). Ump1p was partially stabilized in *pre1-1* cells, as indicated by the increased amount of  $^{35}$ S-labeled Ump1p in *pre1-1*

extracts versus wt extracts that contained the same total amount of TCA-precipitable  $^{35}\text{S}$  (Figure 4A). However, we still observed a significant decrease of pulse-labeled Ump1p upon increasing chase times when the immunoprecipitation was carried out under nondenaturing conditions (Figure 4A). By contrast, when the extracts were treated with 0.4% SDS at 100°C before immunoprecipitation, virtually no decrease of the Ump1p signal during the chase was observed in *pre1-1* cells. These results were consistent with the possibility that the newly formed Ump1p-ha became inaccessible to the anti-ha antibody during the chase because it became trapped within the newly formed *pre1-1* proteasome. If so, Ump1p was expected to be present in fractions from the Superose-6 column that corresponded to the 20S and 26S proteasomes in the *pre1-1* mutant. Indeed, whereas in wt (*PRE1*) cells Ump1p-ha was detected by SDS-PAGE and immunoblotting only in fractions 22 and 23 (corresponding to the ~15S proteasome precursor complex), in the mutant (*pre1-1*) cells Ump1p-ha was detected in fractions 15–23, indicating that it was also present in mature proteasomes (Figure 4B). We assayed the accessibility of Ump1p-ha to anti-ha antibody by immunoprecipitation from fraction 19 (the 20S proteasome) of extracts from the *pre1-1* mutant and from fraction 22 (the ~15S proteasome precursor) of extract from *PRE1* cells. Ump1p-ha could be immunoprecipitated from either fraction 19 or 22 following pretreatment with 0.1% SDS, but not under nondenaturing conditions (Figure 4C), indicating that the ha tag was inaccessible to the anti-ha antibody in both the *pre1-1* proteasome and the precursor complex. However, when otherwise identical assays were carried out with a polyclonal anti-Ump1p antiserum, most of Ump1p could be immunoprecipitated under nondenaturing conditions from fraction 22 (the ~15S proteasome precursor) but was still not immunoprecipitated from fraction 19 (the 20S *pre1-1* proteasome). In contrast, when polyclonal anti-proteasome antibodies were used, Ump1p was precipitated quantitatively along with both complexes (Figure 4C).

These results indicated that a part of Ump1p other than its C terminus is accessible to anti-Ump1p antibodies in the ~15S precursor. However, Ump1p becomes entirely inaccessible (under nondenaturing conditions) upon formation of the 20S proteasome from two ~15S precursors. This interpretation was supported by examining the sensitivity of Ump1p-ha in different complexes to trypsin digestion (Figure 4D). Specifically, Ump1p-ha in the 20S proteasomes (fraction 19) from *pre1-1* cells was completely protected against trypsin. In contrast, Ump1p-ha in the ~15S proteasome precursors was detectably accessible to trypsin. The degradation of Ump1p was incomplete, resulting in a protected fragment of Ump1p-ha that lacked the ~5 kDa N-terminal region but retained the C-terminal ha tag (Figure 4D). Interestingly, in a similar experiment with a strain expressing pro-Pre2p-ha, we observed that, in the Ump1-containing ~15S proteasome precursor complex (fraction 22), pro-Pre2p-ha was shortened by trypsin treatment to yield a product whose electrophoretic mobility was indistinguishable from that of the natural mature Pre2p-ha (Figure 4D). In these experiments, the overall structure of the Ump1p proteasome precursor complex remained intact as judged by native gel analysis of trypsin-treated material (data not shown).

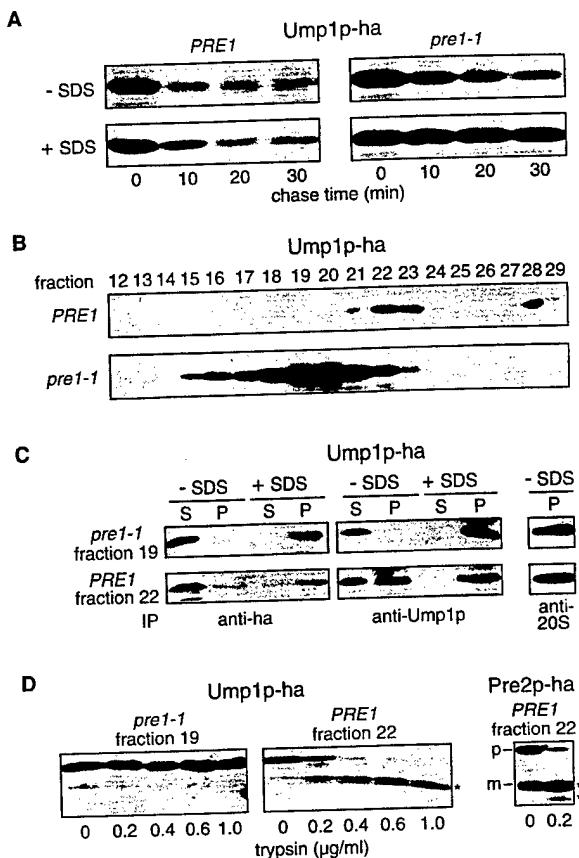


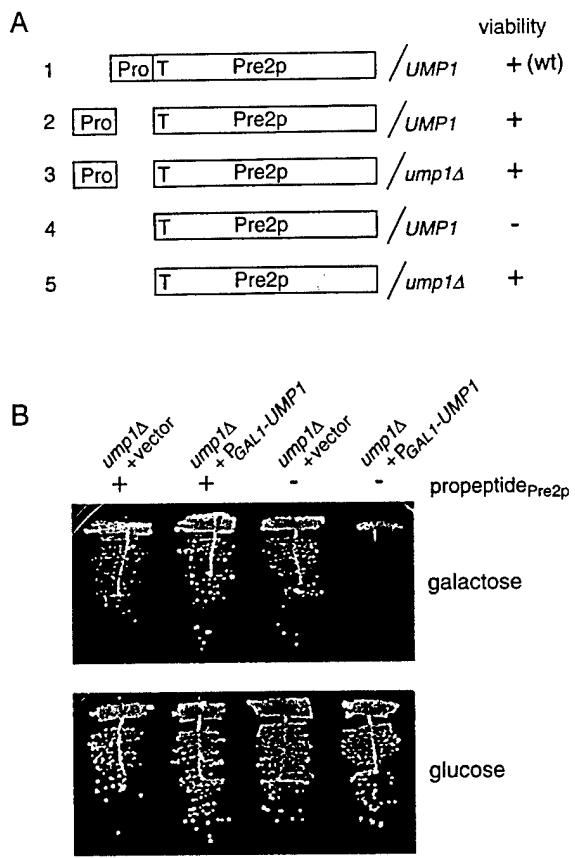
Figure 4. Ump1p Is Stabilized in *pre1-1* Mutants Cells that Are Defective in the Proteasome's Chymotrypsin-like Activity and Persists in 20S Proteasomes

(A) Pulse-chase analysis of Ump1p-ha in wt and *pre1-1* cells. The strains used were JD150 (*PRE1*) and JD151 (*pre1-1*) (Table 1), both expressing Ump1p-ha. The cells were pulse-labeled with  $^{35}\text{S}$ -Met/Cys for 5 min. Extracts were prepared from samples taken at different chase times. Samples were then split in halves. One-half (-SDS) was subjected to immunoprecipitation with anti-ha antibody according to the standard pulse-chase protocol. The other half (+SDS) was adjusted to 0.4% SDS and incubated at 100°C for 5 min, then diluted with extraction buffer to the final concentration of 0.1% SDS prior to immunoprecipitation.

(B–D) SDS-PAGE and anti-ha immunoblot analyses. (B) Detection of Ump1p-ha-containing complexes in extracts fractionated by Superose-6 gel filtration. Strains were the same as in (A). Note the accumulation of Ump1p-ha in fractions containing the 20S and 26S proteasome (fractions 15–21) in *pre1-1*. (C) Immunoblot analyses of immunoprecipitations were carried out with material from Superose-6 fraction 19 of extracts from strain JD151 and fraction 22 of extracts from strain JD150, using the indicated antibodies. Immunoprecipitations were performed without addition of SDS (-SDS) or after boiling in the presence of 0.1% SDS (+SDS). P, precipitates; S, supernatants. (D) Assaying trypsin sensitivity of the same material as in (C), and of the Superose-6 fraction 22, from strain JD138 expressing Pre2p-ha. P, proPre2p-ha; m, mature Pre2p-ha; asterisks mark trypsin digestion products.

#### In the *ump1Δ* Mutant, the Propeptide of Pre2p Is Not Required for Incorporation of Pre2p into the Proteasome

Chen and Hochstrasser (1996) have elegantly demonstrated that the propeptide of Pre2p, if separated from the mature Pre2p, could function in *trans* and thereby



**Figure 5.** The Pre2p Propeptide Is Not Essential in the *ump1Δ* Mutant  
(A) Representation of different genotypes of a set of congenic strains and their effects on cell viability. Strain 1 (JD50-10B) is the wt control. Strain 2 (MHY952) expresses Pre2p/Doa3p-ΔLS with N-terminal Thr (T) as a fusion to ubiquitin (not shown), and the Pre2p/Doa3p propeptide (LS) from two separate plasmids (Chen and Hochstrasser, 1996). Strain 3 (JD160), same as strain 2 but *ump1Δ*. Strain 4 would be derived from strain 2 through the loss of the plasmid encoding Pre2p/Doa3p-LS that operates in *trans*. We were unable to produce isolates that lost this plasmid, confirming the earlier demonstration that the propeptide is essential for cell viability (Chen and Hochstrasser, 1996). Strain 5 (JD163) is derived from strain 3 through the loss of the propeptide-encoding plasmid. Loss of the plasmid did not affect cell viability in the *ump1Δ* background.

(B) Induction of *UMP1* expression in the *ump1Δ* background inhibits the growth of a strain that lacks the Pre2p propeptide. Strains 3 (plus) and 5 (minus) shown in (A) were transformed with pJD429 that expressed *UMP1* from the galactose-inducible, glucose-repressible  $P_{GAL1}$  promoter, or with an empty vector as a control. Transformants selected on glucose media were pregrown on selective media containing raffinose as a carbon source, then streaked onto selective media with 2% galactose or 2% glucose, and grown for 3 days at 30°C.

still allow the incorporation of Pre2p into the 20S proteasome. Under these conditions, expression of the propeptide is essential for cell viability, suggesting that the propeptide, in *cis* or at least in *trans*, is required for the assembly of an active proteasome (Chen and Hochstrasser, 1996).

We confirmed this result using their strain MHY952, in which mature Pre2p and the propeptide are expressed on separate plasmids (Figure 5A). Specifically, under nonselective growth conditions this strain did not yield viable cells that had lost the plasmid expressing the

propeptide. Surprisingly, however, when we constructed and examined a congenic *ump1Δ* derivative of this strain, we noticed that it lost the propeptide-expressing plasmid at a frequency suggesting that this plasmid did not provide a significant growth advantage to cells (data not shown). Indeed, growth rates of cells with the plasmid were indistinguishable from those lacking it (Figure 5B). This result demonstrated that *ump1Δ* is a suppressor of a deletion of the propeptide of Pre2p. If so, expression of *UMP1* in the (*ump1Δ PRE2-Δpro*) background should be incompatible with cell viability. This prediction was confirmed when we examined growth properties of the (*ump1Δ PRE2-Δpro*) strain transformed with a plasmid expressing *UMP1* from the galactose-inducible  $P_{GAL1}$  promoter. On glucose-containing media, the growth rate of this transformant was indistinguishable from that of an otherwise identical transformant carrying the propeptide-expressing plasmid. In contrast, on galactose-containing media, the (*ump1Δ PRE2-Δpro*) mutant containing the  $P_{GAL1}$ -*UMP1* plasmid was unable to grow, whereas the control strains grew (Figure 5B). This result indicated that the propeptide of Pre2p is essential for the Ump1p-assisted maturation of the proteasome but is not essential for (partially) defective maturation of the proteasome that takes place in the absence of Ump1p.

## Discussion

### **Ump1p, a Novel Maturation Factor of the 20S Proteasome**

We describe the discovery of a proteasome maturation factor, termed Ump1p, whose unique properties include a short *in vivo* half-life that is due to its degradation within the newly formed proteasome. Ump1p is required for coordination of the proteasome's physical assembly and enzymatic activation. In addition, the normally essential propeptide of the Pre2p  $\beta$  subunit was found to become nonessential in the absence of Ump1p. We report the following specific results.

### ***ump1Δ* Mutants Are Defective in Ub-Mediated Proteolysis**

They are sensitive to a variety of stresses and accumulate Ub-protein conjugates (Figure 1 and data not shown). All of the observed phenotypes of *ump1Δ* cells are consistent with the conclusion that *ump1* mutants are impaired in proteasome biogenesis and, consequently, in the degradation of ubiquitylated proteins.

### **The UMP1 Gene**

It encodes a polypeptide with a calculated molecular mass of 16.8 kDa. No close sequence homologs of Ump1p were detected in the current databases. However, the presence of a small protein similar in size to Ump1p in preparations of the half-proteasome precursors in mammals (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997) suggests the presence of a functional homolog of Ump1p in the mammalian proteasome maturation pathway.

### ***Ump1p* Is a Component of Proteasome Precursors that Contain Unprocessed $\beta$ Subunits**

The Ump1p-containing complex has a molecular mass of 300–400 kDa, sediments at ~15S, migrates significantly faster in native gels than the 20S proteasome,

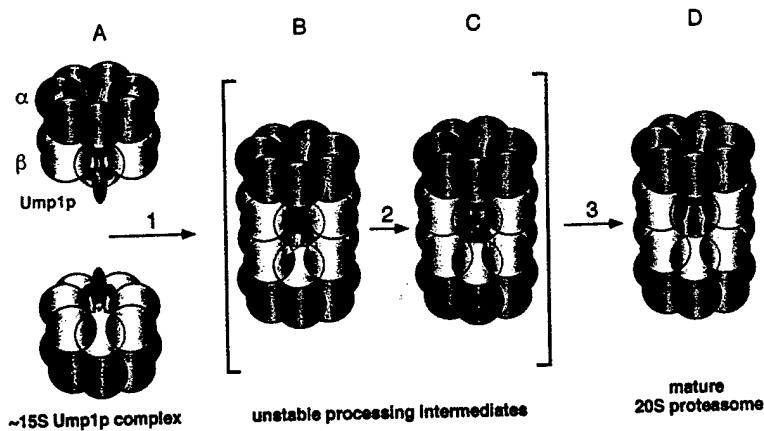


Figure 6. A Model of the Ump1p Function in Proteasome Maturation

Shown is a schematic view of the 20S proteasome and its precursor forms, with the  $\alpha$  and  $\beta$  subunits as blue and green balls, respectively. Three of the  $\beta$  subunits are drawn with extensions that represent propeptides. The  $\beta$  subunits in the front are drawn transparent in order to allow a view into the interior chamber of the proteasome. Structure A is a proteasome precursor complex (half proteasome), characterized by the presence of Ump1p and unprocessed  $\beta$  subunits. In step 1, two of these precursors join to build structure B, a step that leads to conformational or positional shifts of Ump1p and propeptides. Conformational changes of the propeptides trigger their autocatalytic processing (step 2),

and activation of the proteasome's proteolytic activities. This leads to the degradation of the chamber-entrapped Ump1p by the newly formed proteasome (step 3). Only structures A and D are long-lived enough to be detected in wt cells. The findings with the *pre1-1* mutant, in which the degradation of Ump1p is inhibited, suggest the existence of the short-lived intermediates B and C. See text for details.

and has a subunit composition that is highly similar to that of the 20S proteasome, as determined by SDS-PAGE and immunoblot analyses with anti-20S proteasome antibodies (Figure 2 and data not shown). Taken together, these results indicate that the Ump1p is contained within a half-proteasome precursor complex, whose counterpart has been described in the maturation pathway of mammalian proteasomes (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidknecht et al., 1997).

#### **Maturation of Pup1p, Pre2p, and Pre3p Is Strongly Impaired in the *ump1 $\Delta$* Mutant**

Considerably increased amounts of the unprocessed (propeptide-containing) forms of these  $\beta$  subunits were detected in the 20S and 26S fractions from *ump1 $\Delta$*  cells (Figure 3). In addition, the proteasomes assembled in *ump1 $\Delta$*  cells had reduced activity. Thus, Ump1p is required for the correct and efficient maturation of the proteasomes.

**Partially Processed Forms of Pro-Pre2p Are Detected in the Proteasome Precursor Complex and the Proteasome of *ump1 $\Delta$*  Cells but Not in WT Cells**

These incompletely and prematurely processed forms of Pre2p (Figures 3A and 3B) appear to underlie a chymotrypsin-like activity that was associated with the precursor complex in *ump1 $\Delta$*  cells but was absent from the same complex of wt cells (Figure 3C). This was suggested by the observation that this activity was not inhibited by treatment with lactacystin and was absent from cells lacking the Pre2p propeptide (P. C. R. and R. J. D., unpublished data). These findings indicated that Ump1p has a dual role in proteasome maturation. Specifically, Ump1p prevents premature processing of proPre2p in the precursor complex and is also required for the correct maturation of active sites upon assembly of the proteasome (see below).

#### **Ump1p Is Degraded by the Newly Formed Proteasome**

The kinetics of the rapid degradation of Ump1p is similar to the kinetics of processing of Pup1p (Figures 3B and 4A), suggesting that Ump1p is destroyed by the proteasome upon its formation from the two half-proteasome precursors and the accompanying maturation of active

sites (Chen and Hochstrasser, 1996). This conclusion was strongly supported by the observation that Ump1p is significantly stabilized in the *pre1-1* mutant, which is deficient in the proteasome's chymotrypsin-like activity (Figure 4A). Specifically, in contrast to the pattern in wt cells, Ump1p was detectable in the 20S and 26S forms of the *pre1-1* proteasome. Ump1p in these complexes was shielded both from detection by antibodies and from digestion by trypsin (Figures 4B–D). These and related data suggested a model in which Ump1p is enclosed within the 20S proteasome upon its assembly from the two Ump1p-containing half-proteasome precursors (Figure 6). Formation of the 20S structure triggers active site maturation, resulting in the degradation of Ump1p.

#### **The Propeptide of Pre2p Is Not Essential in *ump1 $\Delta$* Cells**

Chen and Hochstrasser (1996) demonstrated that the propeptide of Pre2p is essential for cell viability and that this propeptide can operate in *trans*. They concluded that the propeptide, in addition to rendering proteasome precursors proteolytically inactive, serves a chaperone-like function required for the correct incorporation of Pre2p into the proteasome. One striking result of the present work is that the propeptide of Pre2p becomes dispensable in the *ump1 $\Delta$*  mutant (Figure 5). In fact, the presence of the propeptide, which is essential for viability of wt cells, does not provide a growth advantage to *ump1 $\Delta$*  cells that express it.

#### **On the Functions of Ump1p and $\beta$ Subunit Propeptides in Proteasome Maturation**

Our results suggested a model illustrated in Figure 6. This model is based in part on the idea described by Chen and Hochstrasser (1996)—that the active sites of the proteasome are formed upon its assembly from two half-proteasome precursors, through a juxtaposition of subunits at the interface of the dyad-related halves. The advantage of this mechanism is that autocatalytic maturation of the active sites (through processing of the relevant  $\beta$  subunits) is coupled to the assembly of the proteasome, thus avoiding premature processing of the propeptides of  $\beta$  subunits. This view was supported by

Table 1. Yeast Strains

Strain	Relevant Genotype	Source/Comment
YPH500	<i>MAT<math>\alpha</math> ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52</i>	Sikorski and Hieter, 1989
JDC9-11	<i>MAT<math>\alpha</math> ump1-1</i>	Derivative of YPH500
JD47-13C	<i>MAT<math>\alpha</math> his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52</i>	Dohmen et al., 1995
JD59	<i>MAT<math>\alpha</math> ump1-Δ1::HIS3</i>	Derivative of JD47-13C
JD53	<i>MAT<math>\alpha</math></i>	Derivative of JD47-13C
JD81-1A	<i>MAT<math>\alpha</math> ump1-Δ1::HIS3</i>	Derivative of JD53
JD75	<i>MAT<math>\alpha</math> PRE1-ha::Yiplac211 ump1-Δ1::HIS3</i>	Derivative of JD47-13C
JD126	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128 PRE1-Flag-6His::Yiplac211</i>	Derivative of JD47-13C
JD127	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128 PRE1-ha::Yiplac211</i>	Derivative of JD71
JD129	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128</i>	Derivative of JD127
JD131	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128 PUP1-ha::Yiplac211</i>	Derivative of JD129
JD132	<i>MAT<math>\alpha</math> ump1-Δ1::HIS3 PUP1-ha::Yiplac211</i>	Derivative of JD59
JD133	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128 PRE2-ha::Yiplac211</i>	Derivative of JD129
JD134	<i>MAT<math>\alpha</math> ump1-Δ1::HIS3 PRE2-ha::Yiplac211</i>	Derivative of JD59
JD135	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128 PRE3-ha::Yiplac211</i>	Derivative of JD129
JD136	<i>MAT<math>\alpha</math> ump1-Δ1::HIS3 PRE3-ha::Yiplac211</i>	Derivative of JD59
JD138	<i>MAT<math>\alpha</math> PRE2-ha::Yiplac211</i>	Derivative of JD47-13C
JD139	<i>MAT<math>\alpha</math> PUP1-ha::Yiplac211</i>	Derivative of JD47-13C
BBY45	<i>MAT<math>\alpha</math> his3-Δ200 leu2-3,112 lys2-801 trp1-1 ura3-52</i>	Bartel et al., 1990
JD61	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> ump1-Δ1::HIS3/ump1-Δ2::LEU2</i>	Congenic with BBY45
JD50-10B	<i>MAT<math>\alpha</math> leu2-3,112 his3-Δ200 trp1-1 ura3-52</i>	Congenic with BBY45
MHY952	<i>MAT<math>\alpha</math> leu2-3,112 his3-Δ200 trp1-1 ura3-52 pre2(=doa3)Δ1::HIS3 (YCpUbDOA3ΔLS-His) (YEpDOA3<sub>LS</sub>)</i>	Chen and Hochstrasser, 1996, congenic with BBY45
JD160	<i>MAT<math>\alpha</math> ump1-Δ3 pre2(=doa3)Δ1::HIS3 (YCpUbDOA3ΔLS-His) (YEpDOA3<sub>LS</sub>)</i>	Derivative of MHY952
JD163	<i>MAT<math>\alpha</math> ump1-Δ3 pre2(=doa3)-Δ1::HIS3 (YCpUbDOA3ΔLS-His)</i>	Derivative of JD160, cured of YEpdOAA3 <sub>LS</sub>
WCG4a	<i>MAT<math>\alpha</math> ura3 his3-11 leu2-3,112</i>	Heinemeyer et al., 1993
YHI29/1	<i>MAT<math>\alpha</math> pre1-1 ura3 his3-11 leu2-3,112</i>	Heinemeyer et al., 1993
JD150	<i>MAT<math>\alpha</math> UMP1-ha::Yiplac128</i>	Derivative of WCG4a
JD151	<i>MAT<math>\alpha</math> pre1-1 UMP1-ha::Yiplac128</i>	Derivative of YH129/1

biochemical analyses of the proteasome maturation in mammalian cells (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997).

The results of the present work further support this model. In addition, we discovered that a novel factor, Ump1p, is required for autocatalytic active site maturation and proteasome assembly to occur in a coordinated fashion. A model that accounts for our findings and is consistent with the available evidence is illustrated in Figure 6. In this model, Ump1p has a chaperone-like function required for the efficient processing of the propeptides of  $\beta$  subunits upon the assembly of 20S proteasomes. Specifically, Ump1p interacts with these  $\beta$  subunits, most likely with their propeptides (structure A in Figure 6). Upon formation of the 20S proteasome (Figure 6, step 1), a conformational change of Ump1p, or a change of its position within the complex, would induce a conformational change of the propeptide that results in autocatalytic processing (Figure 6, step 2).

Why is the propeptide of Pre2p essential for proteasome assembly and activation *only* in the presence of Ump1p? We propose that the propeptide of Pre2p is required to induce an alteration of conformation or position of Ump1p upon the assembly of the 20S particle. In this model, the absence of the propeptide would leave Ump1p in a position that is incompatible with the formation of active proteasome. It is possible (and remains to be verified) that this chaperone-like function of a propeptide is unique to Pre2p, as it contains a much longer propeptide (75 residues) than the other  $\beta$  subunits.

In the model of Figure 6, Ump1p is a metabolically

unstable chaperone that is required for the proper (and properly timed) processing of  $\beta$  subunits upon the assembly of the 20S proteasome and that is destroyed within the newly activated proteasome. The unusual mechanics of Ump1p, its noncatalytic mode of action, and its degradation by the protease it helps to activate characterize Ump1p as a novel type of molecular chaperone.

#### Experimental Procedures

##### Yeast Media

Yeast rich (YPD) and synthetic (S) minimal media with 2% dextrose (SD) or 2% galactose (SG) were prepared as described (Dohmen et al., 1995).

##### Isolation of the *UMP1* Gene

*UMP1* was identified using a selection-based screen for mutants in the N-end rule pathway (reviewed by Varshavsky, 1996). *S. cerevisiae* strain YPH500 (Table 1) was transformed with two plasmids: pJD205, which expressed a Ura3p-based N-end rule reporter substrate (Arg-Tpi1p-Ura3p), and pRL2, which expressed Arg- $\beta$ -gal, another N-end rule reporter substrate that can be monitored using X-Gal plate assays (Baker and Varshavsky, 1995). Previous work (Dohmen et al., 1994) has shown that rapid degradation of a Ura3p-based N-end rule substrate renders the cells Ura<sup>-</sup>, whereas mutants in the N-end rule pathway that express the same Ura3p-based reporter are Ura<sup>+</sup>. In this screen, Ura<sup>+</sup> isolates were selected on SD media lacking uracil and were then tested on X-Gal plates to verify that the same isolates were also defective in the degradation of the Arg- $\beta$ -gal N-end rule substrate. One of the mutants thus identified (*ump1-1*) was cured of pJD205 and transformed with a genomic yeast library (Rose et al., 1987). Six transformants yielded plasmids with four overlapping inserts that could restore the ability of cells to degrade Arg- $\beta$ -gal. A 1981 bp BamHI/Sau3A fragment common

to all of the inserts was sequenced and found to contain one complete ORF and two flanking incomplete ORFs. Further mapping, using subcloning and PCR, confirmed that the complete ORF of 445 bp was responsible for the complementation. The *UMP1* sequence (EMBL database accession number: AJ002557) is identical to ORF YBR173C, an ORF subsequently identified by the yeast genome project.

#### Construction of Yeast Strains and Plasmids

Table 1 lists the strains used in this study. To construct *ump1* $\Delta$  alleles, the 1981 bp fragment described above was subcloned into M13mp19. Using single-stranded DNA of the resulting phage, a synthetic oligonucleotide and T4 DNA polymerase, the *UMP1* ORF was precisely deleted and replaced by a BgIII restriction site. The resulting fragment was subcloned into pUC19, and the BgIII site was used to insert fragments containing the *HIS3*, *LEU2*, or the *URA3* gene, the latter one being flanked by two direct repeats of a segment of the *E. coli hisG* gene (Alani et al., 1987). The resulting deletion alleles (*ump1*- $\Delta$ 1::*HIS3*, *ump1*- $\Delta$ 2::*LEU2*, and *ump1*- $\Delta$ 3::*URA3*) were isolated as BamHI fragments, introduced into *S. cerevisiae*, and used to delete the *UMP1* gene (Rothstein, 1991). *ump1* $\Delta$  strains carrying an unmarked *ump1*- $\Delta$ 3 allele (resulting from recombination between the *hisG* repeats) were selected on plates containing 5-fluoroorotic acid (Alani et al., 1987). Construction of chromosomal ORFs that expressed C terminally tagged versions of Ump1p or proteasome subunits (Pre1p, Pre2p, Pre3p, and Pup1p) instead of their wt counterparts was performed as follows. Using primers that contained flanking EcoRI and KpnI sites, 3' portions of the respective genes were amplified by PCR. These sites were then used to insert the amplified fragments into integrative plasmids based on Yplac128 (*LEU2* marked) or Yplac211 (*URA3* marked) (Gietz and Sugino, 1988) that contained sequences encoding epitope tags followed by the terminator sequence of the *CYC1* gene ( $T_{CYC1}$ ). Each of the resulting plasmids was linearized within the coding sequence for targeted integration into the *S. cerevisiae* genome, yielding strains with one copy of the respective gene (fused in-frame to the tag-coding sequence) expressed from its natural promoter and  $T_{CYC1}$ , in addition to a 3' portion of the same gene without promoter. The epitope tags used were double ha ("ha") and Flag-His<sub>6</sub> ("FH") (Dohmen et al., 1995). Plasmid pJD429 (*CEN/URA3*) expressing a *UMP1* cDNA from *PGAL1* was isolated from a cDNA library (Liu et al., 1992) through complementation of the *ump1* $\Delta$  mutation.

#### Pulse-Chase Analyses

R- $\beta$ -gal (Ub-R- $\beta$ -gal) and Ub-P- $\beta$ -gal (Bachmair et al., 1986) were expressed in the *MATa* strains JD47-13C or JD59. *MAT* $\alpha$ 2- $\beta$ -gal was expressed from the plasmid YCp50- $\alpha$ 2<sub>deg</sub>- $\beta$ -gal (Hochstrasser and Varshavsky, 1990) in the *MAT* $\alpha$  strains JD53 and JD81-1A (Table 1). Pulse labeling for 5 min with Redivue Promix [<sup>35</sup>S] (Amersham) followed by a chase and immunoprecipitation with monoclonal anti- $\beta$ -gal (Promega) or anti-ha (16B12, Babco) antibodies were carried out as described by Dohmen et al. (1991). <sup>35</sup>S proteins fractionated by SDS-PAGE were detected by fluorography.

#### Fractionation of Whole-Cell Extracts by Gel Filtration

*S. cerevisiae* were grown at 30°C in YPD or in SD media to OD<sub>600</sub> of 1.4 ± 0.1, harvested at 3000g, washed with cold water, frozen in liquid nitrogen, and stored at -80°C. Cell paste was ground to powder in a mortar in the presence of liquid nitrogen. The extraction buffer was 50 mM Tris-HCl (pH 7.5), 2 mM ATP, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 15% glycerol, used at 2 ml per gram of pelleted yeast cells. After centrifugation at 31,000g for 10 min at 2°C, the supernatant was subjected to a second centrifugation at 60,000g for 30 min, yielding an extract with the protein concentration of ~5 mg/ml. The protein concentration of extracts was equalized in parallel experiments, using the extraction buffer. Using the FPLC system (Pharmacia), 200  $\mu$ l samples of an extract were chromatographed on a Superose-6 column equilibrated with extraction buffer. The flow rate was 0.3 ml/min and fractions of 0.6 ml were collected. The Superose-6 column was calibrated using the following standards: thyroglobulin (669 kDa), ferritin (443 kDa), and bovine serum albumin (66 kDa). Dextran blue was used to monitor the void volume.

#### Assays for Proteolytic Activities with Fluorogenic Peptide Substrates

To determine the chymotrypsin-like activity, 20  $\mu$ l of the protein fraction and 20  $\mu$ l of 0.5 mM succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin in 50 mM Tris-HCl (pH 7.8), 2 mM ATP, 5 mM MgCl<sub>2</sub>, and 1 mM DTT were mixed and incubated for 1 hr at 37°C. The reaction was stopped by addition of 960  $\mu$ l of cold ethanol, and the fluorescence was measured at 440 nm, using the excitation wavelength of 380 nm. The trypsin-like activity and the peptidylglutamyl peptide-hydrolyzing activity were determined with, respectively, Cbz-Ala-Ala-Arg-4-MeO- $\beta$ -naphthylamide and Cbz-Leu-Leu-Glu- $\beta$ -naphthylamide as fluorogenic peptide substrates (Fischer et al., 1994), except that the volume of the protein fraction was 75  $\mu$ l, the reactions were stopped after 270 min at 37°C, and the fluorescence emission was measured at 366 nm, using an excitation wavelength of 420 nm. One unit (U) of a proteolytic activity is defined as 1  $\mu$ mol of the fluorophore produced per min under these conditions.

#### Immunoprecipitation, Electrophoresis, and Immunoblotting

Immunoprecipitations were carried out using either monoclonal anti-tag antibodies (see below), or polyclonal rabbit antisera raised against the yeast 20S proteasome (a gift from K. Tanaka), or against Ump1p-His<sub>6</sub> expressed in *E. coli*. For immunoblotting, the protein samples were boiled for 3 min in the presence of 2% SDS and 0.1 M 2-mercaptoethanol, then subjected to 12% SDS-PAGE, and thereafter transferred onto a PVDF membrane (Millipore) in a wet blot system (BioRad). The blots were incubated with either rabbit anti-ubiquitin (Ramos et al., 1995), or anti-Cim3p (Ghislain et al., 1993) polyclonal antibodies, or 16B12 anti-ha (BabCo), or M2 anti-Flag (Kodak) monoclonal antibodies, and were processed as described (Ramos et al., 1995), except that the initially blotted proteins were visualized using horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgGs, the chemiluminescence blotting substrate detection system from Boehringer Mannheim, and X-ray films. Nondenaturing 4.5% acrylamide gel electrophoresis was performed as described by Hough et al. (1987), and the gels were incubated for 15 min in transfer buffer containing 0.1% SDS before electrotransfer.

#### Affinity Purification of Proteasomes and Related Complexes

The 20S proteasome and its precursors were purified from strain JD126 that expressed Pre1p-FH and Ump1p-ha. Crude extracts were prepared as described above, followed by an exchange of buffer to 50 mM Na-phosphate (pH 8.0), 0.3 M NaCl, using PD-10 columns (Pharmacia). The extract was then incubated with Ni-NTA agarose (Qiagen) in batch for 2 hr at 4°C, followed by washings and elution according to the manufacturer's protocol, except that the (pH 6.0) washing buffer contained 20 mM imidazole. His<sub>6</sub>-tagged proteins were eluted with a step gradient of 100–500 mM imidazole. Active fractions containing a mixture of the mature proteasome and its precursors were pooled, diluted 2-fold with 50 mM Tris-HCl (pH 7.5), and incubated for 2 hr at 4°C in batch with 0.5 ml anti-Flag M2 antibody affinity resin (IBI/Eastman Kodak) that had been equilibrated in TN buffer (0.15 M NaCl, 50 mM Tris-HCl, [pH 7.5]). After the loading, the resin was washed with TN buffer. Flag-tagged proteins were specifically eluted with the Flag epitope peptide (IBI/Eastman Kodak).

#### Acknowledgments

We thank Rohan Baker, Anthony Bretscher, Ricardo Ferreira, Wolfgang Heinemeyer, Wolfgang Hilt, Mark Hochstrasser, Carl Mann, Nancy Kleckner, Keiji Tanaka, and Dieter Wolf for the gifts of plasmids, yeast strains, and antisera; Elisabeth Andrews for assistance in the cloning of *UMP1*, Reiner Stappen for assistance in producing anti-Ump1p antibodies, Robert Kramer (Limberg Druck) for printing the figures, and Niils Johnsson and Ralf Kölking for helpful suggestions. P. C. R., J. H., and R. J. D. are grateful to Cornelis P. Hollenberg for providing lab space and for his support, and to Isabel Fuchs for technical assistance. P. C. R. was supported by a postdoctoral fellowship from Fundação para a Ciência e Tecnologia, Programa Praxis XXI. This work was supported by a grant to R. J. D. from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (0316711), by start-up funding from the Ministerium für

Wissenschaft und Bildung des Landes Nordrhein-Westfalen, and by a grant to A. V. from the National Institutes of Health (GM31530).

Received December 10, 1997; revised January 16, 1998.

## References

- Alani, E., Cao, L., and Kleckner, N. (1987). A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. *Genetics* 116, 541–545.
- Arendt, C.S., and Hochstrasser, M. (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. USA* 94, 7156–7161.
- Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234, 179–186.
- Baker, R.T., and Varshavsky, A. (1995). Yeast N-terminal amidase—a new enzyme and component of the N-end rule pathway. *J. Biol. Chem.* 270, 12065–12074.
- Bartel, B., Wünnigen, I., and Varshavsky, A. (1990). The recognition component of the N-end rule pathway. *EMBO J.* 9, 3179–3189.
- Chen, P., and Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* 86, 961–972.
- Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* 65, 801–847.
- Deveraux, Q., Ustrelli, V., Pickart, C., and Rechsteiner, M. (1994). A 26S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* 269, 7059–7061.
- Dohmen, R.J., Madura, K., Bartel, B., and Varshavsky, A. (1991). The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. *Proc. Natl. Acad. Sci. USA* 88, 7351–7355.
- Dohmen, R.J., Wu, P., and Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* 263, 1273–1276.
- Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. *J. Biol. Chem.* 270, 18099–18109.
- Fischer, M., Hilt, W., Richter-Ruoff, B., Gonen, H., Ciechanover, A., and Wolf, D.H. (1994). The 26S proteasome of the yeast *Saccharomyces cerevisiae*. *FEBS Lett.* 355, 69–75.
- Frentzel, S., Pesold-Hurt, B., Seelig, A., and Kloetzel, P.-M. (1994). 20S proteasomes are assembled via distinct precursor complexes. *J. Mol. Biol.* 236, 975–981.
- Ghislain, M., Udvardy, A., and Mann, C. (1993). *S. cerevisiae* 26S protease mutants arrest cell division in G2/metaphase. *Nature* 366, 358–362.
- Gietz, R.D., and Sugino, A. (1988). New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene* 74, 527–534.
- Groll, M., Ditzel, L., Loewe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463–471.
- Heinemeyer, W., Gruhler, A., Mohrie, V., Mahe, Y., and Wolf, D.H. (1993). *PRE2*, highly homologous to the human major histocompatibility complex-linked *RING10* gene, codes for a yeast proteasome subunit necessary for chymotryptic activity and degradation of ubiquitinated proteins. *J. Biol. Chem.* 268, 5115–5120.
- Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H. (1997). The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* 272, 25200–25209.
- Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* 30, 405–439.
- Hochstrasser, M., and Varshavsky, A. (1990). In vivo degradation of a transcriptional regulator: the yeast α2 repressor. *Cell* 61, 697–708.
- Hochstrasser, M., Papa, F.R., Chen, P., Swaminathan, S., Johnson, P., Stillman, L., Amerik, A.Y., and Li, S.J. (1995). The doa pathway—studies on the functions and mechanisms of ubiquitin-dependent protein degradation in the yeast *Saccharomyces cerevisiae*. *Cold Spring Harb. Symp. Quant. Biol.* 60, 503–513.
- Hough, R., Ratt, G., and Rechsteiner, M. (1987). Purification of two high molecular weight proteases from rabbit reticulocyte lysate. *J. Biol. Chem.* 262, 8303–8313.
- Jentsch, S., and Schlenker, S. (1995). Selective protein degradation—a journey's end within the proteasome. *Cell* 82, 881–884.
- Johnson, E.S., Ma, P.C., Ota, I.M., and Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* 270, 17442–17456.
- Jungmann, J., Reins, H.A., Schobert, C., and Jentsch, S. (1993). Resistance to cadmium mediated by ubiquitin-dependent proteolysis. *Nature* 361, 369–371.
- Liu, H., Krizek, J., and Bretscher, A. (1992). Construction of a GAL1-regulated yeast cDNA expression library and its application to the identification of genes whose overexpression causes lethality in yeast. *Genetics* 132, 665–673.
- Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268, 533–539.
- Lupas, A., Flanagan, J.M., Tamura, T., and Baumeister, W. (1997). Self-compartmentalizing proteases. *Trends Biochem. Sci.* 22, 399–404.
- Nandi, D., Woodward, E., Ginsburg, D.B., and Monaco, J. (1997). Intermediates in the formation of mouse 20S proteasomes: implications for the assembly of precursor β subunits. *EMBO J.* 16, 5363–5375.
- Peters, J.M. (1994). Proteasomes: protein degradation machines of the cell. *Trends Biochem. Sci.* 19, 377–382.
- Ramos, P., Cordeiro, A., Ferreira, R., Ricardo, C., and Teixeira, A. (1995). The presence of ubiquitin-protein conjugates in plant chloroplast lysates is due to cytosolic contamination. *Austr. J. Plant Physiol.* 22, 893–901.
- Rose, M.D., Novick, P., Thomas, J.H., Botstein, D., and Fink, G.R. (1987). A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. *Gene* 60, 237–243.
- Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.* 194, 281–301.
- Schmidtko, G., Schmidt, M., and Kloetzel, P.-M. (1997). Maturation of mammalian 20S proteasome: purification and characterization of 13S and 16S proteasome precursor complexes. *J. Mol. Biol.* 268, 95–106.
- Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R., and Baumeister, W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* 268, 579–582.
- Seemüller, E., Lupas, A., and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. *Nature* 382, 468–470.
- Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122, 19–27.
- Varshavsky, A. (1996). The N-end rule—functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* 93, 12142–12149.
- Varshavsky, A. (1997). The ubiquitin system. *Trends Biochem. Sci.* 22, 383–387.
- Yang, Y., Früh, K., Ahn, K., and Peterson, P.A. (1995). In vivo assembly of the proteasomal complexes, implications for antigen processing. *J. Biol. Chem.* 270, 27687–27694.